

Who wants a poisoned chalice?

Mr George Bush as president should quietly abandon the Strategic Defense Initiative and accept the strict interpretation of the Anti-Ballistic Missile Treaty in return for a deal on strategic arms.

EVEN the people who have spent the past year or so hoping that Mr George Bush would not now be the next president of the United States have been heaping sympathy on him since last Tuesday's election, which is a very curious state of affairs. To judge from the judicious head-shaking that is going on, it might be thought that Bush is about to inherit a bed of nails, not one of the most interesting jobs still open to competition. It is not even the case that he will be taking over at an especially bad time: the US economy, having grown at a steady 3 per cent a year for the past few years, can put even more resources than previously at his disposal, while the international outlook is more settled, and more cheerful, than for 35 years. Is it possible that the sympathy being lavished on Bush stems not so much from calculations of the tricks a Congress controlled by his political opponents may play, as from envy masquerading as sympathy?

The new president will have just three urgent tasks on his desk on 20 January. The first is to do something more effective than the present government has done to control the budget deficit (see *Nature* 336, 2; 1988); that is important not only to escape the thrall of the international money markets in a fashion hinted at by last week's 3 per cent fall of the dollar against the mark and yen, but also to provide the emollient increase of prosperity without which change cannot be contemplated. The second task is to come to an arrangement about the defence of Europe not only with the Soviet Union, and Mr Mikhail Gorbachev in particular (there is to be a summit meeting next summer), but with Western Europe as well. Finally, Bush has to show that he means business in promising to make the United States a "gentler" place; it will not be possible to accomplish very much in four short years.

Luckily for Bush, there is an obvious way in which he can put the first two problems together and make them vanish as if they were a particle and its anti-particle; cut defence spending so as to make the budget balance better, but make an offsetting deal with the Soviet Union to provide security of a different kind. It is not as if the budget deficit is that far beyond control. Saving something like \$30,000 million a year in each of the next three years would get rid of it altogether (allowing for the increase of tax revenues that supposed steady growth would bring), but this sum is not very different from the amounts of money the Department of Defense and the Congress have been arguing about during the closing stages of the budget process in the past few rough years.

Opportunity

Bush's opportunity is nicely dramatized by the case of the Strategic Defense Initiative (SDI), which now costs about \$5,000 million a year and which has been for the past three years one of the chief obstacles, at the Geneva negotiations, to an agreement on the limitation of strategic missiles. Dropping SDI after President Reagan has left office would probably not save the whole of what SDI now costs, but there are much bigger savings to be made elsewhere in the Pentagon's strategic military budget. Indeed, the turnabout could even be accomplished by sleight of hand, without having ever to confess that SDI could never be the leak-proof shield against missile attack which it has

been advertised as being since 13 March 1982. It would probably suffice if the United States agreed to abide by the strict interpretation of the Anti-Ballistic Missile Treaty (which would forbid testing of active components of the system) for, say, a period of ten years. (Gorbachev wants a longer commitment, but that is unreasonable.) That way, the new president could never be accused of having "thrown away" a powerful negotiating card, for the option to resume development would remain.

Trouble

This, of course, is where Bush's troubles will begin. His own constituents, in Europe as well as the United States, are divided in their opinions of the pace with which a rapprochement with the Soviet Union can safely proceed. At one extreme is what seems to be the view of the West German government: that Gorbachev's arrival, and his evident good sense, provide an historic opportunity for reaching a civilized understanding on European security. At the other there remains a profound suspicion of the Soviet Union's motives and military objectives.

In the United States, two particular issues hamper the emergence of clear policies; the legacy of decades of rhetoric (some of which may have been justifiable, but which was rarely justified) about the "evil empire" in the East, and the conviction that Western Europe should shoulder a larger share of the burden of its own defence. Mrs Margaret Thatcher, apparently among the agnostics on the significance of recent Soviet change, is nevertheless right to have been pushing this week for a meeting of North Atlantic Treaty Organization politicians next summer to thrash these questions out. The danger ahead is that what seems to be the prudent course — cautiously to wait and see how affairs turn out in the Soviet Union — may be calamitous. Those who think otherwise should heed Andrei Sakharov's warning last week that the West should be less concerned with the risks in helping *perestroika* along than with the risk that *perestroika* might fail (see page 191).

Sadly, Bush will find himself in a box on US policy towards the Soviet Union and, for that matter, towards the rest of the world, which should point him quickly towards the third urgent item on his January agenda. While US industries are widely admired for their nimbleness in changing course, US public policy is too often the prisoner of compromise between opposing views most easily recognized by their stridency. Perhaps for constitutional reasons, US governments find it more difficult than most others to formulate moderate policies and then stick by them (which does not always save them from being timorous). A moderate plan to make a deal on strategic missiles, followed by a vigorous attempt to sell it to the 54 per cent of Americans who voted Republican on Tuesday last week, would help to create some of the conditions in which 'gentleness' could have half a chance. To go the whole way, Bush will have to pay more attention to the educational system than he or his predecessor seem inclined to do. The two issues may seem at the opposite ends of the political spectrum, but they are closely connected. The United States remains as vigorous a melting-pot as ever for people and ideas, but needs better ways of matching what it wants with what it can hope to get. □



Japan asks for more

University chemists in Japan want more money. They might have asked for even more.

Is the chronically parlous condition of Japanese universities an anachronism or part of the foundation of Japan's economic success? That is a question on many Western politicians' minds since the 1960s, when the helplessness of Japanese academics in the face of scarce support was accidentally drawn to wider attention by the student troubles of the times; some politicians, as British academics know too well, have responded as if they had discovered in Japan a way of making academic scientists productive: make university life so uncomfortable that they prefer to work in industry.

To be fair, the scene is not uniform. While most university departments and the professorial research teams inhabiting them are badly housed and equipped, some are quite the opposite — often because of a senior person's links with industry through formal contracts (from the revenues of which the university, as with grants from government sources, usually takes a share). It is as serious that the assistants attached to a productive professor, essentially his postdoctoral helpers, may languish in that state for years, badly paid and demoralized by the need to wait for dead men's shoes to become vacant. Yet so rigid are the rules that some able people shoulder two parallel teaching loads to acquire extra help.

The most striking feature of the report on page 197 of conditions in chemistry departments in Japanese universities is its source — not a committee appointed by a government agency seeking to sustain a predetermined policy, but the independent Chemical Society of Japan, which has taken care to poll university chemistry departments throughout the country before giving its opinions of what should now be done. The bill that the society would like the government to pick up is substantial but in no sense outrageous. Perhaps nuclear magnetic resonance (NMR) machines have been used too readily as a yardstick for comparison and a justification for the appearance of the begging-bowls, but they are, as the world knows, literally indispensable means of characterizing compounds, new or old.

Whether the Japanese government will respond as the chemists would wish is another matter. Chemists are not the only paupers on the scene, for one thing. And the government will say, as on previous occasions, that it has no money, which is also true. (The Japanese budget deficit is larger, as a proportion of GNP, than that of the United States, but its economic consequences are unimportant because it is financed, at low rates of interest, by Japanese citizens saving for their retirement, which comes early.) What it should appreciate is that sooner or later it will have to find some, if only to hide its shame. Meanwhile, academics might prod the government (and some of their colleagues) towards a more enlightened organization for the public universities as a whole.

The *koza* system, for example, needs urgently to be reformed. While many of these teaching and research teams are excellent means of training graduate students, others are but little empires in which able people lose enthusiasm and drive. A looser structure could benefit from a more generous system of competitive grants, and is almost a precondition for its success.

The management of the large national universities also cries out for change. As often elsewhere, the administration is powerful, but not often as remote as in it can be at some Japanese universities. But the greatest need of all is that there should be a simpler way in which the Japanese contribution to science can be more smoothly made a part of what is happening elsewhere, not simply in the West but in Asia. That, of course, is a tall order: a few NMR machines will be easier to find.

None of this implies that Japan is indifferent to the problem of its downtrodden academic community. During the past decade

in particular, there has been a commendable effort to provide national facilities accessible to all, while national laboratories such as those at Okazaki carry the traditions of basic research into a novel and well-heeled environment. But these developments are not a substitute for enabling academic scientists to pursue research effectively in their own laboratories, and in a way that helps to provide their own students with the stimulation that is just as necessary in Japan as elsewhere. And what the chemists are now saying might just as easily have been said by people from other disciplines. □

Not by appointment

British architects are in the pillory again, but the real villains go scot-free.

PRINCE Charles, who turned forty this week, and who may yet be the King of England (or, strictly, of the United Kingdom), has so offended modern British architects with his complaints about the buildings they design that the profession may neglect the lessons it should be learning, and which are not those apparently on offer. The row has been simmering ever since he called one intended building a "carbuncle". Now, in 75 minutes of near prime-time television, he has laid about the architects with even greater vigour; one building (still a plan) he likened to a "thirties wireless" (otherwise "radio receiver"). The charge is that British architects scorn the people who will live or work in them, and that they have designed monstrosities on the successive tides of changing fashion without consideration for the built environment of the sites falling into their clutches. To a very large extent, the complaint is accurate, but it is a pity that architects are to shoulder the whole blame.

Societies get the buildings they deserve both morally and in the literal sense that they commission architects and eventually pay for what is built. For some considerable time, British clients (even when spending public money) have lacked the qualities that would allow them to discharge their parts in their implicit compacts with their architects. A sense of good aesthetic taste, valuable though that may be because it is so often wanting, may be less important than their lack of clarity on the question of how they wish to live and work. How else does it come about that the most common complaint against British architects is that they are forever "playing God" with other people's lives? Clients with an inkling of what they wanted would hardly let that happen to them on such an important matter.

British clients are also scandalously money-conscious. They expect that their architects will not merely ensure that the plumbing is silent but that the buildings they eventually own will be recognized as works of art by others even if they are themselves uninterested, but expect this to be done on a percentage fee. But clients (or their bankers) do have strong views on critical parameters of building design that might affect the commercial values. Influenced, no doubt, by knowing that many sixteenth-century buildings are still used, every client sees his investment as a permanency — and is abetted by planners with regulatory authority who will not easily allow buildings, however ill-distinguished, to be replaced by others.

Prince Charles, regrettably, may have helped confirm potential clients in their cowardice. It is one thing to be ridiculed by friends and acquaintances, but quite another to risk being told off by a monarch in waiting. So there will now be a temptation for clients to be even more cautious and for architects to grope towards a rose-cottaged style they hope may win a plug on some future television programme. But the plain truth is that there is no such thing as a characteristic contemporary style, the post-modernists notwithstanding. Most architectural traditions, and the British in particular, are more in need of experiment and the diversity it brings than of a new orthodoxy. That is why Prince Charles has exerted his influence in the wrong direction. □

Sakharov visits US to launch international group

- *Perestroika* under domestic threat
- Sakharov's first visit to the West

Boston

SOVIET physicist and dissident Andrei Sakharov embarked on his first visit to the West last week to help launch a new international foundation. This week he is scheduled to meet President Reagan at the White House, and on Wednesday is due to meet for the first time Edward Teller at a reception in Washington.

In his remarks to a large gathering of scientists and reporters at the American Academy of Arts and Sciences in Cambridge, Massachusetts, the 67-year-old recipient of the 1975 Nobel Peace Prize spoke positively about *perestroika* but also warned of domestic threats which could cripple newly established Soviet freedoms.

Perestroika is not "theatre for the West", Sakharov said, but is an "extraordinarily serious" process which is now in an "acute phase". The Western world, said Sakharov, should not fear *perestroika*'s success, but rather its failure. Such a failure, Sakharov warned, would lead to "external expansion" by the Soviet Union, a "strengthening of repressive organs" in the country.

Sakharov's visit marked the official start in the United States of the International Foundation for the Survival and Development of Humanity, the first international, non-profit organization to be based in the Soviet Union. The foundation, which has established offices in Moscow, Stockholm and Washington, DC, plans to make grants for collaborative international projects in the areas of arms control, environmental protection, international development, energy and human rights, and, according to Sakharov, to "take a role in the global problems now standing in front of humanity".

Along with prominent Soviet scientists Yevgenii P. Velikhov and Roald Sagdeev, Sakharov will spend approximately three weeks in the United States to help promote the fledgling international organization and to convene its first board meeting with members from eighteen nations. The organization, which has preliminary funding from the Soviet Union and from Armand Hammer, 90-year-old chairman and Chief Executive Officer of Occidental Petroleum, plans to begin funding proposals soon after 1 January 1989.

Velikhov, the chairman of the foundation's executive committee, and vice-president of the Soviet Academy of Sciences, told the gathering that problems of the environment, disarmament and the biosphere "do not belong to any one

country" and that this realization formed the "groundwork" for establishing the foundation. Sakharov's presence, he said, was an indication of the Soviet Union's current commitment to the "freedom of ideas".

Sakharov was allowed to leave the Soviet Union under a provision of a law passed one month ago which allows international foundations to exist within the Soviet Union and includes a provision permitting Soviet board members to attend foundation meetings.

Jerome Weisner, vice-chairman of the foundation's executive committee and president emeritus of Massachusetts Institute of Technology, stressed the need for many nations to involve themselves in the foundation's work. Other foundation board members include former US Secretary of Defense Robert McNamara, John Sculley, Chief Executive Officer of Apple Computers, and representatives from Japan, China and many European countries.

At the press conference to launch the foundation, Sakharov, in his first public appearance in the West, fielded a wide range of questions from reporters in which he condemned the Soviet Union's cessation of withdrawal of their troops from Afghanistan and called for UN intervention in that conflict. He also expressed his views that space-based defence initiatives like the US Strategic Defense Initiative are "very destabilizing" and "completely senseless from an economic point of view". Many of his remarks, however, were devoted to his concerns about "extremely dangerous" recent laws in the Soviet Union that threatened *perestroika* by limiting people's freedom of association and demonstration and by centralizing authority within the Soviet ministry of internal affairs, creating, he said, a situation of virtual "martial law".

In addition, Sakharov, who was banished to the closed city of Gorky in January 1980 after calling for the withdrawal of Soviet troops from Afghanistan, said that it is "vital to succeed in freeing all prisoners of conscience" and stressed that many in the Soviet Union were still being held for their religious and political beliefs. Among such prisoners still being held captive in the Soviet Union, according to Sakharov: Vazif Meylanov, Yuri Badzyo, Levko Lukyanenko, in internal exile in Siberia, and Boris Mityashiv, Mikhail Kukobaka and Mikhail Alekseev, in labour camps or prisons. Seth Shulman

Science gap growing

THE gap between Soviet and Western science is still growing, Dr Nikolai Laverov, one of the five new vice-presidents of the Soviet Academy of Sciences, says in *Izvestiya*. The gap is particularly apparent, he says, in the lack of state-of-the-art instruments for academy institutes, in the mathematical modelling of geophysical and geochemical processes, in the study of mineral ores and in oceanography and sea-floor studies. Laverov was elected last month as vice-president responsible for Earth sciences, in place of Dr Aleksandr Yanshin, who resigned the post under the academy's new compulsory retirement policy. V. R.

New museum director

THE British Museum (Natural History) has a new director. Dr Neil Robert Chalmers has taken over from Dr Ron Hedley, who had held the post for 12 years and has now retired. Chalmers studied zoology at the universities of Oxford and Cambridge and has worked at the Makerere University College, Uganda, the National Primate Centre, Kenya, and the Open University in Britain. The director's appointment is for an initial period of five years. C. McG.

Latest NIH institute

THE National Institute of Deafness and Communication Disorders became the thirteenth and newest institute of the National Institutes of Health (NIH) earlier this month. NIH associate director Jay Moskowitz will be acting director of the institute until a permanent director is found. The institute is authorized to spend up to \$96.1 million in its first year, with funds administered through the National Institute of Neurological and Communicative Disorders and Stroke at the discretion of the Secretary of the Department of Health and Human Services. J. P.

Backing basic research

WITH no relief from the Australian government, Mr Barry Jones, the Minister for Science, Customs and Small Business, is trying to establish a privately backed investment fund for basic research. Jones had chief responsibility for science until a major ministerial reorganization last year.

Jones hopes that 10–20 large corporations will contribute up to A\$100 million each to the fund over ten years and has approached several industry leaders for discussions. Tax exemptions could provide an incentive to invest. Jones has been highly critical of Australia's failure to make long-term investments in new technology although he recognizes that the prime minister and government have to deal with more pressing economic problems. "... This is not the time to engage them in discussion about the long-term future..." he says. C. M.

Orbiting nuclear reactors still a problem for astronomers

Berkeley

OPPONENTS of nuclear reactors in Earth orbit have tried to bring gamma-ray astronomers on board, with a promise that a ban on orbiting reactors could end their problems with spurious background noise. Joel Primack, a theoretical physicist at the University of California at Santa Cruz and a member of the Federation of American Scientists (FAS), which opposes the use of nuclear reactors in Earth orbit, recently organized a session on orbiting reactors at a workshop on high-resolution gamma-ray cosmology in Los Angeles.

The gamma-ray astronomers may be reluctant to join forces for the ban, but they have nevertheless begun to discuss difficulties they have been experiencing for years with spurious background radiation apparently caused by orbiting Soviet reactors.

Members of gamma-ray detection projects including the High Energy Astronomical Observatory and the Solar Maximum Mission (Solar Max) have for

years noticed spurious bursts of background radiation that complicate their measurements, which under the best of circumstances involve a very low signal to noise ratio.

German scientist Eric Rieger, a collaborator on the Solar Max, was the first to identify the radiation bursts as a discrete component of the background noise. The bursts, which seem to originate from orbiting nuclear reactors on Soviet RORSAT reconnaissance satellites, come in two varieties: gamma rays detected directly from the unshielded reactors, and clusters of positrons that remain trapped in the Earth's magnetic field, and annihilate with a burst of radiation on contact with a detector.

The US government has apparently known of the phenomenon for several years, but has kept the information classified, says Primack. Last month, Primack saw an unclassified National Aeronautics and Space Administration (NASA) memo, dated August 1988, that described the radiation bursts and warned that they

could jeopardize data collection by NASA's planned Gamma Ray Observatory. Primack, who was an invited speaker at the Los Angeles workshop earlier this month, requested time for a second presentation on the proposal for a reactor ban, supported by the FAS and the Committee of Soviet Scientists Against the Nuclear Threat (see *Nature* 334, 465; 1988).

After Primack's presentation, Solar Max scientist Edward Chupp, of the University of New Hampshire, spoke about problems the bursts had caused for the Solar Maximum Mission, and James Kurfess of the Naval Research Laboratory confirmed that the bursts had been recognized for some time, and their potential influence on the Gamma Ray Observatory were being studied.

Primack says many of the scientists in the audience acknowledged that they too had been troubled by the background, but had never understood its source.

Solar Max project scientist David Forrest, of the University of New Hampshire, says the positron annihilations generated much excitement when they were first identified by Rieger, shortly after the Solar Maximum satellite was launched. But he says it soon became clear from the periodicity of the events that the positrons were from a man-made rather than astronomical source.

Computer programs were developed to identify and cancel the spurious signals, but studying weak gamma-ray sources such as Supernova 1987a was nevertheless made much more difficult. In addition, Forrest says the events have become dramatically more frequent in the past year, for unknown reasons, occurring as often as once per orbit.

The NASA memo that originally alerted Primack to the problem warns that the bursts could jeopardize the Gamma Ray Observatory by setting off a trigger meant to initiate observation of astronomical sources of gamma-rays. Once activated, the sensor will be unable to respond to gamma-ray bursts from astrophysical events for up to one orbit. The memo asks Gamma Ray Observatory scientists to study the problem and seek means to avert it.

While banning nuclear satellites would be one solution to the problem — and Primack is quick to point out that the 'fleet' of powerful nuclear reactors planned as part of the Strategic Defense Initiative may be devastating to future, more sensitive gamma-ray sensors — the astronomers seem to believe that ways can be found to get around the problem. "Gamma-ray astronomers don't tend to have much influence in controlling the outside environment", says Forrest. The most they can hope for, he says, is to develop a means to recognize and subtract the spurious signals. **Marcia Barinaga**

Lasker prizes shared by three researchers

Washington

THE Albert Lasker Medical Research Awards for 1988 go to a clinical researcher for his work on drug addiction and two molecular biologists for their studies of the role on RNA in living cells.

The medical research award is shared by Thomas Cech of the University of Colorado and Phillip Sharp of the Massachusetts Institute of Technology. Cech's work has shown that polyribonucleotides can act as enzymes, a discovery that Frank Westheimer wrote "is revolutionary in its direct impact on the theory of catalysis, in its promise for future discovery, and its implications with respect to prebiotic chemistry" (*Nature* 319, 534; 1986). In his work, Sharp showed how genes can be split up by non-coding sequences of DNA,

and how these non-coding sequences are excised once they are translated to RNA and before they are translated into proteins. Vincent Dole wins the clinical award for demonstrating that heroin addiction can be chemically controlled with methadone.

A fourth award for public service goes to Senator Lowell Weicker of Connecticut for his "consistent and successful championship of federal funding for a wide range of health care programs" during his 18 years in the Senate.

Each prizewinner receives a \$15,000 honorarium, as well as the knowledge that the Lasker award is frequently the precursor of even greater accolades. Last year, for example, the medical research award went to Susumu Tonegawa, who subsequently won the Nobel prize. Joseph Palca



Lasker winners Thomas Cech (left), Phillip Sharp (centre) and Vincent Dole (right).

New directions, new faces for science following US elections

Washington, Berkeley & Boston

THE transition between the Reagan and Bush administrations has begun with the announcement that Mr James Baker will be the next Secretary of State. Beyond that, the Bush transition team has made no detailed pronouncements about the direction in which the new administration will go, or whether there will be changes in support for science. But the results of elections for membership of Congress ensure that several key figures in science policy will disappear from the national scene.

President-elect George Bush seems set to continue President Ronald Reagan's strong support for 'big science' projects, such as the Superconducting Super Collider (see page 195) and the space programme. He has supported the plans initiated by the Reagan Administration to double the budget of the National Science Foundation, but appears less zealous than Reagan in his support for the Strategic Defense Initiative.

The Democrat-controlled Senate faced by Bush will begin its next session lacking three members who played key roles in guiding science policy. Democrat William Proxmire of Wisconsin retires from the Senate, where he had been chairman of the subcommittee overseeing the budget of the National Science Foundation.

Proxmire was known for his efforts to staunch the flow of government dollars to wasteful projects. He awarded a 'Golden Fleece' award to particularly egregious examples of waste, although the awards often rankled with researchers who felt they were unfairly attacked.

Florida Senator Lawton Chiles (Democrat) announced earlier this year that he would not run for re-election. Chiles had been chairman of the appropriations subcommittee responsible for the budget of the National Institutes of Health. Also missing will be Republican Lowell Weicker of Connecticut who lost an extremely close race to Democrat Joe Lieberman. Weicker had been an active supporter of biomedical research, and had a strong interest in ocean research.

Congress will take note of public feeling expressed directly at state ballots on a number of science-related issues. Californians split their votes on two measures involving AIDS. One, that would have required notification to public health officials of anyone testing positive for antibodies to the AIDS virus, was rejected by 66 per cent of the voters, although polls taken several months ago had indicated it would pass (see *Nature* 335, 386; 1988). The proposition received the support of California governor George Deukmejian, but was opposed by most Californian

medical associations.

A second initiative calling for AIDS testing of accused criminals that received far less publicity passed by a 62 to 38 per cent margin. It allows the victims of sex crimes, as well as police officers, firemen and emergency medical personnel who are bitten, spat on or otherwise exposed to bodily fluids during an assault, to request that the accused assailant be tested for human immunodeficiency virus.

Despite a well-financed campaign by the tobacco lobby, Californians voted to increase the tax on a package by 25 cents. In Oregon, a measure that would have

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The president-elect in less than zealous mood?
banned smoking in most indoor public places went down to defeat.

In Massachusetts, voters soundly rejected a referendum question that would have shut the state's two nuclear power plants (see *Nature* 336, 3; 1988). The vote was 32 per cent for the measure, 68 per cent against. The measure's supporters blame the scale of their defeat on the fact that the opposition, organized by regional utility companies, raised over \$7 million in campaign funds, more than twenty times the money raised by the measure's proponents.

Three states passed measures that will make abortions more difficult to obtain. Those in Colorado and Michigan forbid the use of state funds to pay for abortions. In Arkansas, voters adopted the so-called Unborn Child Amendment, which could have the effect of forbidding all abortions in the state. The act's full impact will be felt if the Supreme Court reverses its landmark decision that made abortions legal, as it may do during its current term.

Reactors on order despite accusations

Washington

In the wake of a series of accidents and charges of mismanagement at existing facilities, plans to construct two new reactors to produce tritium needed for nuclear weapons are "moving forward on a fast track", a Department of Energy (DoE) official said last week. But even under the best of circumstances it will be "ten years before the reactors will be on-line". Troy Wade, acting assistant secretary of the DoE's defence programmes, said that work should begin within two years on construction of a new heavy-water reactor and a modular high-temperature gas reactor capable of producing tritium. The new facilities will replace ageing reactors at the Savannah River Plant in South Carolina, none of which is currently operating.

Wade said DoE expects the Department of Defense to provide the \$6,800 million needed for the reactor programme, and predicted there would be some interagency bickering over the issue. The programme must also get congressional approval. Wade said that the United States is, for the first time, living off its "working reserve" of tritium and if production does not start up again soon the situation will become critical.

Lisa Perlman

Neural networks programme to start

Washington

A MAJOR new research programme to develop artificial neural network computers will be launched by the US Defense Advanced Research Projects Agency (DARPA) this month. The programme begins by providing almost everything that was asked for in a report commissioned by DARPA and published three months ago.

For the first 17 months of the programme, \$33 million will be available for research aimed at assessing the advantages of neural-network computers in complex information processing and control systems, developing new modular hardware to make it easier to build large neural-network computers and advancing theory to determine the limitations of neural-network systems.

That will take the programme to the end of fiscal year 1990, beyond which DARPA is not yet able to commit itself. The report to DARPA asked that initial funding of \$40 million be followed by increases which would provide a total of \$400 million for a programme running until 1995. Evaluation of progress will centre on the success of neural-network computers at three tasks of great interest to the Pentagon: automatic target recognition, continuous speech recognition and sonar signal identification.

Alun Anderson

Medical journal in gun-ban challenge

Washington

IN Maryland the health consequences of hand-gun ownership became a referendum issue and ended up testing the much-debated publication policies of the *New England Journal of Medicine*. The referendum had its roots in a bill adopted by the Maryland legislature to restrict the availability of so-called Saturday-night specials, cheap hand-guns often used in assaults. The National Rifle Association supported a vigorous assault on the new law, obtaining enough signatures to force a state-wide vote.

On 8 November, the day of the vote, the influential *Washington Post* newspaper decided to break a *New England Journal of Medicine* embargo and give advance publicity to an article showing that hand-gun controls appear to reduce the frequency of urban violent crimes. The article, by John Henry Sloan *et al.*, appearing in the 10 November issue, argues that the smaller number of homicides and aggravated assaults in Vancouver relative to neighbouring Seattle, Washington is related to Canada's strict gun-control laws.

Arnold Relman, editor-in-chief of the *Journal*, who earlier barred Reuters from receiving advance copies after the news agency broke an embargo declined to comment on the *Washington Post's* decision. But he said "our position on the embargo has never been absolute. On rare occasions when there is the perception by us or others that the information in an article is so important that it ought to be released immediately, we have lifted the embargo."

Whether the *Washington Post* article had a significant impact is unknown. But the following day voters threw out the challenge to the gun ban. It was the National Rifle Association's first defeat in a state referendum.

Joseph Palca

Refuges threaten endangered species

Washington

US NATIONAL Wildlife Refuges are killing the creatures they are supposed to be saving, according to a report released two weeks ago by the Wilderness Society. The federal government has appropriated \$51.7 million for the upkeep and preservation of these refuges, but the society claims that commercial development and agricultural runoff containing pesticides and heavy metals are ruining wildlife refuges. Kesterson National Waterfowl Refuge in California became a *cause célèbre* when toxins in surrounding waters killed birds that landed there. The society says that officials at Kesterson have to spend much of their energy trying to scare wildlife away.

Lisa L. Lyles

Global warming becomes an international political issue

Hamburg

THE discussion about global warming due to a 'greenhouse effect' could become one of the major political issues of the next century, according to delegates at the *Climate and Development* conference in Hamburg last week. The debate is being fuelled by the results of studies on the effects of temperature on resources such as water and energy, and on tourism, agriculture and other industries. If global temperature changes are to be controlled, concerted international action on an international level will be necessary, but political wrangling could obstruct a consensus.

Discussions at the conference centred around means of reducing emissions of greenhouse gases, but Mikhail Budyko of the Soviet state committee for hydrometeorology outlined the potential benefits of greenhouse gas emissions for agriculture in barren areas such as the tundra. His argument was received like "swearing in the church", said one delegate. But it is an argument that is sure to emerge again as politicians attempt to assess what action is needed to tackle climate change.

A plan for the development of government policies was formulated at the meeting. But more research is needed before governments will implement any policies; research not just into the scientific facts but into impacts on the economy and society of each nation.

One of the problems hindering such studies, said Martin Parry of the University of Birmingham, is the absence of a robust model with which to analyse the economic costs and benefits of adapting to

climatic change. In a report for the Organisation for Economic Cooperation and Development (OECD), which will be published next month, he recommends that a methodology be formulated for member countries. At the beginning of next year, the OECD will begin a study of the greenhouse effect and its implications for future energy use and economic policies — and aims to formulate methods for cost-benefit analysis and risk assessment.

Also embarking on a major study of the socio-economic impact of climate change is the inter-governmental panel set up by the World Meteorological Organisation and the United Nations Environment Programme. When the panel met last week for the first time it set a tight schedule for the study: the programmes of each of three working groups will be reviewed in about three months, with full reports expected in 18 months. The results of both studies will be presented at the second world climate conference of UNEP and the WMO in 1990. If an international law of the air is considered necessary, a preliminary draft of the law could be available by 1992.

In the meantime, the conference timetable is filling up for researchers, politicians and environmentalists concerned with the greenhouse effect. One delegate said he knew of about 15 such conferences next year. The "travelling circus" of the greenhouse debate has begun, said another last week. If each one tries to better its predecessors, then delegates might be calling for a 100 per cent reduction in CO₂ emissions before the year is out.

Christine McGourty

US-Soviet Solar System collaboration

Washington

CLOSER cooperation between the United States and the Soviet Union in the exploration of the Solar System is promised in an agreement signed in Washington last week by senior representatives from both sides. Some joint work, including exchange of data and the participation of scientists from one country in the missions of the other, is already under way, but the new agreement opens up the possibility of Soviet spacecraft carrying US instruments, and vice versa. Academician Valery Barsukov, head of the Soviet delegation, announced that all future Soviet missions to Mars, planned and expected, would now be open to international cooperation.

The agreement stems from a general programme of joint discussions set into motion with the signing in April 1987 of a US-Soviet treaty on cooperation in space science. The joint working group on Solar System exploration has met twice since

then, and both sides say they have been pleased by the speed with which cooperative projects have moved. Data from US Viking missions to Mars are being used by the Soviets to select possible Mars landing sites, and the Soviet Phobos-2 craft is being tracked by the US Deep Space Network as it journeys to Mars.

No specific plans have yet been made for missions in which the two countries will share spacecraft facilities or even build them together, but one possibility mentioned by Barsukov is a sample-return mission to Mars. Although cooperation in space missions between the two countries has a long history, the new agreement marks the first time that detailed technical specifications of spacecraft will be exchanged. Manned exploration of space has not been discussed so far, but there was general optimism that there would also eventually be cooperation in that area.

David Lindley

Superconducting Super Collider decision made at last

Washington

THE town of Waxahachie, 25 miles south of Dallas, Texas, was named last week as the intended home of the Superconducting Super Collider (SSC), the 52-mile-circumference proton accelerator that the US Department of Energy (DoE) hopes to build. Secretary of Energy John Herrington's announcement officially ended eighteen months of fierce and expensive competition for the SSC between a number of states, but provoked a chorus of protests from the losers that the outcome had been politically rigged.

Since the selection of seven candidate sites from an initial entry of 43 late last year (see *Nature* 331, 2; 1988), a DoE task force has visited the sites, assessed geological, environmental and demographic data, heard testimony from the proponents and estimated for each location the 'life-cycle' cost of building the SSC and running

account from the site selection task force, the Illinois proposal had one important failing. Adjoining the Fermi National Accelerator Laboratory, the SSC site goes under what is now a moderately urban area, and packets of land would have had to be obtained from over 3,000 individual land-owners.

The DoE task force was not convinced that the state would have been able to deliver the land to the federal government on schedule, especially as there was concerted local opposition to the SSC. This problem merited the only 'poor' rating handed out by the task force.

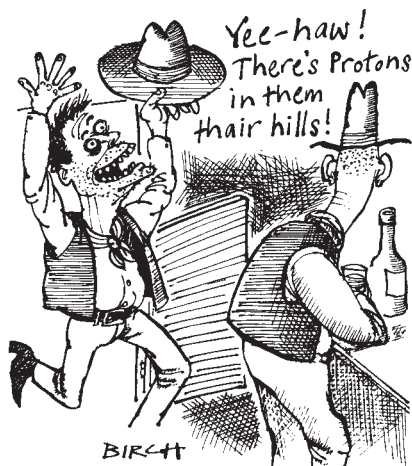
The advantage of Illinois, according to its supporters, was that Fermilab's Tevatron, a four-mile-long proton ring, could have been used as a feeder for the main SSC ring, eliminating a substantial fraction of the construction cost. But the life-cycle cost estimated by the task force was \$10,400 million, on the unproved assumption that the Tevatron could be plugged into the SSC with only minor changes, and rose to \$10,900 million if major changes were needed. By contrast,

the cost at the Texas site was estimated at \$10,800 million.

Among the other sites, none had any serious disadvantages, but neither did they have the overall strengths of the Texas proposal. The Waxahachie site has easy access to the Dallas-Fort Worth metropolitan area, presents no special environmental problems, and the local community has supported the SSC bid.

Although Governor James Martin of North Carolina and Senator Dennis DeConcini of Arizona have indicated that they want to see some kind of inquiry into the site selection, they have not produced any specific objections to the technical ratings on which the choice was made. DoE will probably be able to defend itself against charges of political bias.

For Texas politicians, the selection of Waxahachie was cause for celebration. Among those attending the Herrington press conference were Jim Wright, Speaker of the House of Representatives, and Senators Lloyd Bentsen and Phil Gramm, along with two local congressmen. Texan George Bush announced his support for the project during the presidential campaign. All their efforts are likely to be needed if money to build the SSC is to be extracted from an increasingly parsimonious Congress. **David Lindley**



it for 25 years. Each site was rated under six general headings, the most important being its geological suitability for the tunnels that will house the main ring of magnets. The Texas site was deemed 'outstanding' in four of the six headings, and good in the other two, a ranking that put it clearly ahead of the other six proposals.

Immediate and vehement protest came from Senator Alan Dixon (Democrat) of Illinois, who called the choice "deeply disappointing" and implied that the announcement had been delayed until after the presidential election in order not to offend voters in Illinois. But Herrington said that representatives from all the seven states had asked him not to announce the decision until after the election, and that neither President Ronald Reagan nor President-elect George Bush was told of the outcome until election day.

The Illinois and Texas bids were generally thought to be the top contenders, but, according to the published

International Peace Week of Scientists

Washington

Thanks to the United Nations, this year might be the last that the International Peace Week of Scientists passes unnoticed by most Western scientists. On 8 November, the UN Special Political Committee resolved that an 'International Week of Science and Peace' be proclaimed to take place each year during the week in which 11 November falls. Member states are urged to encourage scientific institutions to hold lectures and special activities on science and peace and to report to the UN Secretary General.

The resolution was introduced by Carlos Gutierrez, the Costa Rican Ambassador to the United Nations, on behalf of Costa Rica's president, Oscar Arias Sanchez. Final approval by the General Assembly is expected this month. UN action builds on an unofficial Science and Peace Week, now in its third year of operation, which was put together by Hendrik Bramhoff, a 33-year-old mathematician at the University of Hamburg in West Germany.

Bramhoff was inspired to set up the Science and Peace Week early in the 1980s when he realized how important scientific opinion could be in arms negotiations, particularly on such technical issues as the policing of test-ban treaties. He found various activities already under way, many of them loosely centred on 11 November, Veteran's Day in the United States. He

helped to set up an International Coordinating Committee for Science and Peace Weeks held in 1986-88 and found backers for a UN resolution.

Bramhoff is now bowing out to concentrate on research and is passing leadership to David Krieger, president of a peace foundation based in Santa Barbara. Krieger inherits an organization whose biggest obstacle to growth may be its lack of international balance. In building up the Science and Peace Week, Bramhoff said he "worked with those willing to cooperate" and extended invitations to sponsor the week to many scientific organizations, both East and West.

The result is that Science and Peace Week is endorsed in the Eastern bloc by government-sponsored organizations such as the Hungarian Peace Council and the Soviet Peace Committee whereas in the United States and Europe it attracts the support of pressure groups critical of their governments. Bramhoff hopes that with UN support, Science and Peace Week will attract wider attention.

Armed with the resolution, researchers will have a better chance of persuading university presidents to back lectures and other activities. That, he says, is particularly important in developing countries where the links between science and peace have yet to be publicly debated.

Alun Anderson

Novel sponsor makes a debut in European technology research

Verona

THE European Institute of Technology (EIT), an eight-month-old consortium of multinational corporations, made an intriguing debut in biotechnology research funding at its second symposium, in Verona, last week. An enigma to both public and private sectors as well as raising eyebrows at the European Commission, EIT nevertheless received more than 850 research proposals from universities in response to an April call for tenders. Even though it had less than 1 million ECU (European currency units; 1 ECU = £0.83) to hand out, EIT whet appetites with the prospect of 250,000 ECU per year, for three years, and the unusual offer to pay short-listed applicants 15,000 ECU to prepare a more detailed proposal.

EIT is the brain-child of its president, John Marcum, whose tour as director for science, technology and industry of the Organisation for Economic Cooperation and Development (OECD) in 1984-87 ruffled a few diplomatic feathers. The institute has three principal aims. First, it hopes to foster partnerships between industry and university in the fields of biotechnology, information and materials technology by providing grants for pre-competitive research considered likely to yield patentable innovations. Second, EIT promises its industrial members access to a network of leading public-sector researchers. Finally, the institute hopes to offer training courses in technology management through workshops, 30-day courses and a one-year master's programme.

The idea arose after a round-table of European industrialists abandoned a plan to create a bricks-and-mortar European technology institute. Instead, EIT was set up simply as an office in the Paris business neighbourhood, La Defence, with a staff of less than ten and constituted as a non-profit-making educational establishment.

So far, five corporations have paid the 200,000-ECU annual subscription to join the consortium: EniChem, IBM Europe, Montedison, Philips and AT&T, giving EIT a start-up capital of 1 million ECU. With such small resources, explained Marcum, EIT needs to be highly selective, both in terms of the research areas it will sponsor and in the quality of specific projects. "We are an elitist organization", he says, adding that many industry-university collaborations fail because 'second-rate' researchers are chosen from each sector. To solicit interest from leading university researchers — and to create the promised networks — EIT decided to offer the prospect of large grants. But only between five and ten of the 850 proposals will finally be approved

and at rates of an average 100,000 ECU per year for three years, making the 250,000 ECU advertised grants look to some like an ingenious but effective bootstrapping ploy.

Marcum is nevertheless confident that EIT's initiative will pay dividends. Any patents resulting from research it has sponsored will belong to EIT and royalties will be shared in proportion to the sums invested. "We will have to be tough about this", said Marcum. "In the past, universities undercharged, now they're overcharging." Having collaborated in generating the new ideas, sponsoring companies will be free to compete to develop marketable products, or seek partnerships through EUREKA, Europe's market-led programme to get competitive advanced technology ideas into production. Meanwhile, EIT expects to make enough from royalties eventually to be self-sufficient, at least in its education venture.

Superficially similar to EUREKA, EIT has taken care to avoid overlap, by including EC, national government and public-sector representatives on its three Technology Expert Panels. Klaus Draxler, present chairman of EUREKA's group of national programme coordinators (NPCs), who was in Verona to find out more about EIT for the Community, con-

firmed that the two organisations could be mutually beneficial. EIT essentially scours university research groups for budding new concepts (top-down), while EUREKA aims to bring existing ideas, without limitations of domain, to fruition (bottom-up).

Draxler was more concerned that EIT's research programme would clash with the Community's own pre-competitive advanced technology projects, such as ESPRIT, RACE and BRITE. There is also a risk that EIT's education and training programme could conflict with COMETT, which aims to provide Europe-wide training for innovation and development, application and management of new technologies. But he thinks EIT's notion of networks is a good idea. "It could help everyone, including EUREKA. We don't know how fruitful collaborations really come about. It's a matter of contacts. Companies don't have university contacts and so they don't know whom they can work with."

In order to have any lasting influence, EIT will now have to strengthen its industrial membership. Whether the enthusiastic response to EIT's first advertisement is repeated or not is now less important in this respect, as it will have flushed out most of the best ideas and identified who should be in the 'networks'. Meanwhile, 16 of the original applicants (six of them from the United Kingdom) have received small grants to have a crack at the major awards, to be given out next March.

Peter Coles

Finding a punishment to fit the crime

Berkeley

A FORMER chancellor and professor of British history at the University of California, Santa Barbara, who found the campus chancellor's house too unpleasant to occupy, may soon be rubbing elbows with illiterate criminals, teaching them to read as part of his punishment for embezzlement of university funds.

Robert Huttenback became chancellor of the Santa Barbara campus in 1977. In 1979, he and his family moved out of the house provided them by the university, giving as reasons the lack of a neighbourhood environment, a high noise level and the "institutional nature" of the house. The family moved to an expensive Santa Barbara neighbourhood, and received an annual housing allowance from the university for mortgage, maintenance, insurance and property taxes. But Huttenback and his wife also spent university money on home improvements such as a new kitchen and house furnishings.

Huttenback resigned as chancellor in July 1986 after a university investigation concluded that he had inappropriately spent \$174,000 on his home. The investiga-

tion found no wilful intent to deceive the university, and the matter was considered settled by Huttenback's resignation and agreement to repay the money.

But a separate state investigation last year concluded that the inappropriate expenditures totalled \$271,000, and Huttenback and his wife were charged with embezzlement and evasion of taxes. They were convicted in July 1988; and in September they were fined a total of \$70,300 and Huttenback was ordered to perform 1,000 hours of community service, developing a programme to teach functionally illiterate criminals to read, as part of the Santa Barbara city adult literacy programme.

Huttenback is said to have claimed that he believed his expenditures were justified, because his home was used for university functions. He has been suspended from his position as full professor of British history, while the university reviews the case under the faculty code of conduct and decides what action to take. The first University of California chancellor to be convicted of a felony, Huttenback has filed an appeal. Marcia Barinaga

Japan's research predicament revealed in society's survey

Tokyo

JAPAN's university researchers are worse off than they were 20 years ago, according to a report just released by the Chemical Society of Japan. Research funds for university researchers have failed to keep pace with inflation, and there is a chronic shortage of space, equipment and technicians which is getting worse every year. To rectify the situation, the report calls on the government to invest an extra ¥1,000 million per year in the university research system.

Outlays for research and development in Japan have soared 20-fold since 1965, making the nation second only to the United States in total research expenditure. But nearly all of the increase has been in private industry. The government's share of research spending has dropped from 30 to 20 per cent and university researchers have become much poorer than their cousins in private industry. According to the society's survey, the average researcher in industry now receives the equivalent of ¥20 million in research funds per year (on a 1980 yen basis), 46 per cent up in real terms on 1965, while a university researcher has to make do with ¥8 million per year, 16 per cent less than in 1965.

The real situation is worse than these figures suggest. University researchers receive government funds from two main sources: competitive research grants from the Ministry of Education, Science and Culture and 'regular' research funds from their university which are automatically allotted to each department or teaching unit (called *koza*). The latter funds are awarded according to the nature of the *koza*'s research (experimental or theoretical) and teaching responsibilities (the number of graduate students, for example).

Division of funds

About half of these regular funds are held back to cover university administration costs and the remainder are divided up among the members of the *koza* according to rank (a *koza* is typically headed by one professor, under whom there is one assistant professor and two research assistants).

Koza research funds (excluding university administration costs) have decreased 50 per cent in real terms between 1970 and 1985. Competitive research grants, on the other hand, have trebled during the same period. But as there are only a limited number of large grants many *koza* are suffering from a chronic lack of funds. And, according to Akio Yamamoto of Tokyo Institute of Technology, who heads the

committee that compiled the report, the "average" university researcher receives only about ¥300,000 (\$2,400) per year in discretionary research funds.

Lack of funds is not the only problem. In the 20-year period from 1965, the ratio of technical staff to researchers has plummeted from one to one to less than one to three. The ratio in industry has also decreased but still stands close to one to one.

As a result, many young university researchers spend their time repairing equipment and carrying out manual tasks rather than doing productive research.

Many departments also lack new equipment. For example, the report points out that in the United States, chemistry departments typically have several superconducting NMR (nuclear magnetic resonance) machines but in Japan only 3 out of 10 departments have such machines. The situation is particularly bad in new and provincial universities which do not have the same support as the old 'imperial' universities, such as Tokyo and Kyoto.

The picture is not all gloom. Funds for facilities for 'big science' (accelerator physics, fusion and space research) have trebled between 1975 and 1985. And the research institutes for joint university research, such as the High Energy Physics Laboratory in Tsukuba, the Institute of Space and Astronautical Science in Tokyo and the national research laboratories in Okazaki boast world-class research facilities. But these are the exception rather than the rule.

A questionnaire sent out by the society to 262 chemistry department heads received a 70 per cent response, and nearly all those replying (99 per cent) said that facilities in their department were "insufficient" to some degree. Thirty-five per cent said laboratory space is "desperately inadequate". According to Yamamoto, university regulations for laboratory space were made thirty years ago before there were graduate students, foreign researchers or large equipment and the rules only allow for "test tubes and benches".

Last year, however, saw a remarkable improvement in the situation in some laboratories. As a result of Western pressure, Japan ploughed several million million yen into the domestic economy and about ¥40,000 million (\$325 million) of this was allotted to the upgrade of university facilities. The sudden influx of new funds was like a "divine wind", according to Yamamoto. Professors were "flabbergasted" by the unexpected opportunity to buy new equipment. Many chemistry departments bought superconducting

NMR machines. But some departments simply did not have space for new equipment.

The report calls on the government to take the following measures: competitive research grants from the ministry of education should be doubled from ¥50,000 million per year to ¥100,000 million; "regular" *koza* research funds should be increased by ¥50,000 million per year; and funds for equipment and maintenance should be increased by about ¥40,000 million per year. In addition, the number of postdoctoral fellowships for foreign (that is, Western) researchers should be increased, the budget for books, periodicals and computer networks expanded, and more space made available for laboratories.

Unprecedented increase

Such increases in research funds would be unprecedented but are "not so large", says Yamamoto, when compared, for example, with recent increases in Japan's overseas development aid. And in the present political climate, with Japan under pressure from the United States and other Western countries to improve its basic research system, the "wind is not so unfavourable" for this kind of change, he says.

But money may not be the only problem with Japan's universities. The *koza* system has recently come under severe criticism by Japanese Nobel prizewinner Susumu Tonegawa, now at the Massachusetts Institute of Technology. He says the rigid hierarchy within Japan's university departments destroys creativity. Young researchers are stifled by the university research system which demands subservience to elders. Tonegawa left Japan in the 1960s and he advises "bright" young Japanese researchers to do the same.

The rigid *koza* system has also been attacked frequently for being wasteful of resources. Each department acts like a little independent "kingdom" with its own budget and facilities. As a result, there is tremendous duplication of equipment and facilities and very little interaction between research groups in related fields.

Yamamoto, however, thinks the *koza* system has advantages over the open and competitive research system in the United States. He says that although Japan's university research system does not produce many Nobel prizewinners, it does provide "shelter" for young researchers and yields a steady stream of good basic research.

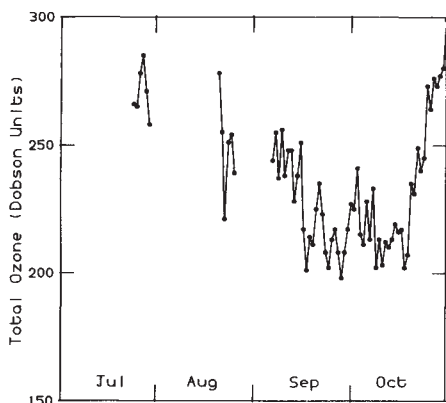
He points to statistics in the report that show that the number of publications from Japanese researchers in two important chemical journals, the *Journal of Organic Chemistry* and the *Journal of the American Chemical Society*, increased three- to four-fold between 1970 and 1985 and now accounts for close to 10 per cent of all contributions.

David Swinbanks

How deep is an 'ozone hole'?

SIR—We wish to comment on the figures for Antarctic ozone depletion quoted by David Lindley (*Nature* 335, 657; 1988), and based on TOMS (total ozone mapping spectrometer) measurements from the Nimbus-7 satellite.

TOMS maps show that the core of the polar stratospheric vortex, in which depletion is most marked, varies rapidly in shape and position, making it difficult to select any one parameter as a measure of the depth of the 'ozone hole'. The figures given by your correspondent refer to September alone (TOMS cannot retrieve ozone values from the polar cap in July and August). The figure below shows, for comparison, provisional ozone amounts



Ozone values at Halley in 1988.

reported to us by our station at Halley Bay. A rough guide to the severity of depletion is given by the difference between values in July and August, and the least values seen in early October, corresponding to a depletion of about 26 per cent. This is similar to that recorded at Halley Bay in 1984 (ref. 1), but less than in 1985, 1986 and 1987.

The vortex has been centred over West Antarctica since early September, and has been elliptical rather than circular (the wave-number one component has been larger than usual, and relatively steady; the wave-number two component has also been large, and has undergone rapid phase progression). Warm air, and high ozone amounts, were established over East Antarctica in late September, and the vortex is now being displaced towards the Atlantic in the course of what appears to be the final warming. The cold core is now passing over Halley Bay. It still contains ozone-poor air; the increase in column ozone seen in late October arises mainly from transport at higher altitudes, above 20 km.

If the strong subsidence seen earlier over East Antarctica continues into West Antarctica as the vortex moves away, the depletion in 1989 could be very deep. The subsided air, rich in ozone and inorganic chlorine compounds, and poor in chloro-

fluorocarbons (CFCs) and nitric oxide (N_2O), will become in the ensuing months next year's polar vortex. Ozone amounts will decay photochemically during the summer, but the inorganic chlorine will persist. The altitude structure of the CFC and N_2O distributions seen in the Airborne Antarctic Ozone Experiment last year will be imposed during the spin-up of the vortex. However, the subsidence in the winter is weak, as may be inferred by the absence of significant changes in ozone amount, and the total amount of inorganic chlorine will remain essentially unaltered.

The extent to which this inorganic chlorine will yield active chlorine in the spring of 1989 will, of course, depend on the temperatures and amounts of polar stratospheric clouds which then prevail. Looking back, the weak depletion of 1988 was determined, at least in part, by the late final warming of 1987, and the weak subsidence which accompanied it. Inorganic chlorine which had subsided earlier into the warm belt was diluted before moving across Antarctica to develop into the 1988 vortex.

Mark Schoeberl, according to Lindley, attributes much of the year-to-year variation in the depth of the 'ozone hole' to the quasi-biennial oscillation (QBO) of the tropical stratosphere. This seems premature. While it is easy to see that the QBO can modulate planetary wave propagation, it seems doubtful whether it controls planetary wave generation, essentially a tropospheric process. It is not unlikely that the strong variations seen in the past few years are largely fortuitous, and that a phase change, independent of the QBO, might well appear.

J. C. FARMAN
B. G. GARDINER
J. D. SHANKLIN

British Antarctic Survey,
High Cross, Madingley Road,
Cambridge CB3 0ET, UK

1. Farman, J.C., Gardiner, B.G. & Shanklin, J.D. *Nature* 315, 207-210 (1985).

Industrial disease?

SIR—The news item "Changing relationship seen in new corporate-university ties" (*Nature* 335, 106; 1988), misses the real reasons for this trend.

Funding is not the issue: rather the issues are research and development costs and patenting and licensing of innovations and inventions. In 1985, the federal government, industry, and academic institutions supplied \$52,003 million, \$49,863 million, and \$2,286 million in R & D funds, respectively (1982 dollars) (*Statistical Abstract of the US: 1988*). Yet they performed \$12,698 million, \$76,956 million

and \$9,439 million in R & D, respectively (1982 dollars). Both industry and academic institutions receive large transfer payments from the federal government in the form of contracts. Despite industry's windfall, problems with intellectual property rights are endemic in government-industry science projects. Patenting and licensing is limited in these joint ventures, reducing profit potential.

Also, in 1985 for example, the average full-time equivalent cost per scientist or engineer was \$140,500, whereas the average graduate student made between \$10,000 and \$15,000 per year. Industry, in an effort to reduce costs, decided that it was better to have graduate students perform the R & D, in the form of PhD theses, at nearly one-tenth the cost per scientist. Even five graduate students were cheaper than one private scientist. Bonuses for industry from joint ventures with academic institutions are greater bargaining power and fewer intellectual property restrictions. This position is much more profitable than that with government.

These facts seem to explain very nicely the growing number of industrial-academic cooperation agreements. As an MIT graduate student, I too fear the loss of academic freedom, as well as the exploitation of graduate students at relatively low wages, with no or few proprietary rights.

DAVID WAGGER
29 Park Drive, Apt 1,
Boston, Massachusetts 02215, USA

Embryo definitions

SIR—Acceptable scientific terminology must aim to be universal and objective. Its integrity must not be compromised for administrative or legal convenience, or to evade ethical considerations. Nonetheless Braude *et al.*¹ have repeated their presumption in employing once again the term "preembryo" as a scientific descriptive, even though their own earlier work had demonstrated the arbitrary nature of the "pre"-fix². It is therefore necessary to re-emphasize that the term was synthesized in the United Kingdom precisely to avoid possible legal restrictions in that jurisdiction on research on human embryos at their earliest stages of development. This lawyer's stratagem cannot constitute a scientific definition. Neither can a voluntary coalition of interested parties assert itself as a 'regulating authority' on ethical issues. Intellectual rigour with precision and objectivity of definition must be maintained, if necessary by the exercise of refereeing and editorial responsibility.

A.J. McEvoy
Institut de Chimie Physique,
EPFL-Ecublens,
CH-1015 Lausanne, Switzerland

1. Braude P., Bolton V. & Moore S. *Nature* 335, 392 (1988).
2. Braude P., Bolton V. & Moore S. *Nature* 332, 459 (1988).

Theoreticians thrive among dunes

The Institute for Theoretical Physics at Santa Barbara has won itself an international reputation in just a decade, but the trick will not easily be repeated.

Santa Barbara, California

THE Institute for Theoretical Physics here seems confident of an even more illustrious future than its immediate, but short, past. Just ten years old, the small permanent staff has had the luxury of interviewing architects during the past few weeks, with the intention of adding to its third request for a further five years of support from the National Science Foundation the cost of a building, which would be shared equally by the University of California at Santa Barbara. The signals from Washington are said to be encouraging; the Gramm-Rudmann Act might have been passed on another planet.

There is more than mere official encouragement to report, but the flattery of imitation as well. The present director of the centre, Dr Robert Schrieffer (the "S" of the BCS theory of classical superconductivity who was calumniously said, in *Nature* two years ago, to have taken a job in Wall Street) says that there have been serious inquiries about the reasons for Santa Barbara's success from people hoping to follow it in Europe. What follows is meant as a thumbnail explanation.

Part of the reason must be the setting, which everybody acknowledges to be a magnet for people from elsewhere. President Ronald Reagan's ranch is in the hills twenty miles north, near the Vandenburg Air Force Base, the back-up shuttle site. The institute is at present housed on the sixth floor of a building at the centre of the university campus established a quarter of a century ago. Although on the edge of the ocean, the campus huddles in the dunes as if afraid of the sea, but the new two-storey building will look directly over the Pacific.

The present members of the institute are anxious to emphasize the degree to which their success stems from their way of working, which in turn stems from the enthusiasm of the then much smaller university for a competition launched in 1968 by the National Science Foundation, which at the outset offered a subvention of \$1 million a year. Part of the calculation then was that the university's own physics department could only profit from the presence of a distinguished, but constantly changing, group of theoreticians. The calculation has been correct; the university's own theoretical physics group is now comparable in size and distinction with that at the institute.

The guiding principle is change. The permanent senior staff numbers just six,

including the director and his deputy and a dozen or 15 postdoctoral workers. But the bread and butter of the institute's work consists of a series of overlapping scientific programmes, perhaps best described as workshops, which last for between 6 and 12 months and each bring 15 or so people to the site at any time. (But a programme on high- T_c superconductivity due to begin next January and to last for six months is so heavily oversubscribed that it is not clear where everybody will sit.) The period is long enough for participants to get serious work done during their stay, as well as to embark on collaborations with other visitors.

Schrieffer (whose five-year term as director will be up in time for him to be an ordinary participant in the superconductivity programme) insists on the care with which the topics of these studies are chosen. The objective, he says, is to find out what the "community" would like to spend six months or more upon, which he emphasizes is consistent with Santa Barbara's role as a national resource. The procedure is that topics are ultimately chosen by the rotating membership of the institute's external advisory board, itself a roster of the working great and good. The trick, Schrieffer says, is to find an able and enthusiastic putative director for each topic at an early stage, giving him a licence to use persuasion and temptation to tell who will be willing to come, for how long, if the board should sanction the proposal.

So far, this process has led the institute to tackle the difficult problems of physics head-on rather than to seek easy popularity in softer fields. The topic nearest to biology, for example, was a programme on "Spin glasses, computation and neural networks" in 1986. Field theory in all its ramification (quantum gravity included) seems to be the central theme, but there is a sustained interest in the cosmological connection, while condensed-matter physics is strengthening its hold on the institute's attention, no doubt because of Schrieffer's own interest and the presence as a permanent member of J.S. Langer, tempted here from the University of Pennsylvania in 1982 and best known for his treatment of the thermodynamics of aggregation in the days before computer simulation became the rage it is now.

While the counting of published papers and their citations is hardly a sure guide to the quality of people's achievements, perhaps the most vivid proof of this institute's

success is the remarkable frequency with which its address turns up in the better physics journals. But the permanent members and the small band of resident postdoctoral fellows have the peculiar advantage of seeing the whole world (or what matters in it) pass along their corridor. "Just stay here long enough" according to one of them, "and you'll have met everybody". He was referring, of course, to the community of theoretical physics.

Santa Barbara is not, of course, the only place with similar objectives. Perhaps the best-known is the Bohr Institute in Copenhagen, now supported under the banner of NORDITA by all Scandinavian governments. The *Institut des Hautes Etudes* near Paris is another. Santa Barbara would say that they are more like ivory towers than it would wish to be. The closest model of this way of working is the International Institute of Theoretical Physics at Trieste, founded by Abdus Salam at about the same time. The new Indian institute at Pune, directed by J.S. Narlikar, will follow a similar pattern but in the narrower field of cosmology, relativity and matter.

What can be learned from the Santa Barbara experience by those with ambitions to set up similar institutes elsewhere? Theoretical physics, being portable, is an obvious field in which to look for success, but it is clear that this institute owes a great deal to the hinterland on which it has been able to draw. It is difficult to see that it could have done as well if the United States had not been as well supplied with theoreticians.

The other crucial consideration is that the budget (which here now just exceeds \$2 million a year) should be large and flexible enough to help with travel and lodging costs. This is notoriously the most difficult kind of money to come by, especially on a regular but untied basis. Some of Santa Barbara's support now comes from the National Aeronautics and Space Administration, but Schrieffer is grateful that it is laundered through the National Science Foundation's accounts. And even then, the permanent members of the institute find it necessary to supplement the institute's funds by means of the usual round of grant application. The moral is that, while \$2 million a year is a sizeable sum, it is not enough to buy a reputation in theoretical physics even in a halcyon climate such as this. **John Maddox**

Cell division

Kinetochores on the move

Tim Mitchison and Anthony Hyman

An important part of the history of mitosis research has been the significance of exceptions. Students of spindle function have long sought out and described aberrant and peculiar situations, from the sex univalents of Dietz¹ to the experimentally retarded monopoles of Mazia². Such exceptions, natural or induced, have acted as important experimental perturbations. However, mitosis with unreplicated genomes (MUGs), as reported by Brinkley *et al.* on page 251 of this issue³, may be the weirdest yet.

Schlegel and Pardee first reported⁴ that when mammalian tissue-culture cells blocked at the beginning of S-phase (before DNA replication) are treated with caffeine, normal cell-cycle controls are over-ridden. The treated cells start to alternate rather rapidly between interphase and mitotic states irrespective of DNA replication, in a manner reminiscent of the cell cycles of early *Xenopus* embryos⁵. Brinkley *et al.*³ have now examined the structural events of mitosis in cells treated with caffeine and hydroxyurea which go through mitosis with unreplicated genomes. They show that not only are mitotic spindles formed, but that they go ahead and perform the full range of mitotic functions, with disastrous consequences for the unreplicated genome.

Segregation of sister chromatids at mitosis is mediated by kinetochores, specialized regions of the chromosome which attach to spindle microtubules. Normally, cells duplicate their kinetochores some time before mitosis. This is

thought to involve replication of centromeric DNA (DNA acts as the site of kinetochore formation), followed by assembly of a new kinetochore on the centromeric DNA. The duplication of the kinetochore structure has been seen in mammalian cells using human autoantibodies that recognize proteins unique to the structure. Single kinetochore spots seen by immunofluorescence turn into pairs of spots some time in G₂, the cell-cycle stage after DNA replication⁶. As the cell enters a normal mitosis, each kinetochore from the pair becomes attached to opposite spindle poles, probably by capturing microtubules nucleated from the poles. Alignment of chromosomes on the centre of the spindle, the metaphase plate, is thought to involve the action of opposing polewards forces on the kinetochore pair. Chromosome segregation at anaphase is thought to be triggered by splitting between the kinetochore pair, allowing the individual kinetochores (and thus the sister chromatids) to be pulled towards the spindle poles (see figure).

In the MUGs, the microtubule system seems to sense nothing amiss. A spindle that looks normal forms, to which the kinetochores of the unreplicated chromosomes attach. During this process, the forces generated by the spindle pull the kinetochores free from the bulk of the chromatin, presumably breaking DNA strands in the process. These free kinetochores, which include the proteins recognized by the human autoantibody⁷

and a small amount of chromatin, then go on to execute a full range of mitotic movements. This includes alignment of kinetochores equidistant from the two poles, and subsequently their segregation into daughter cells.

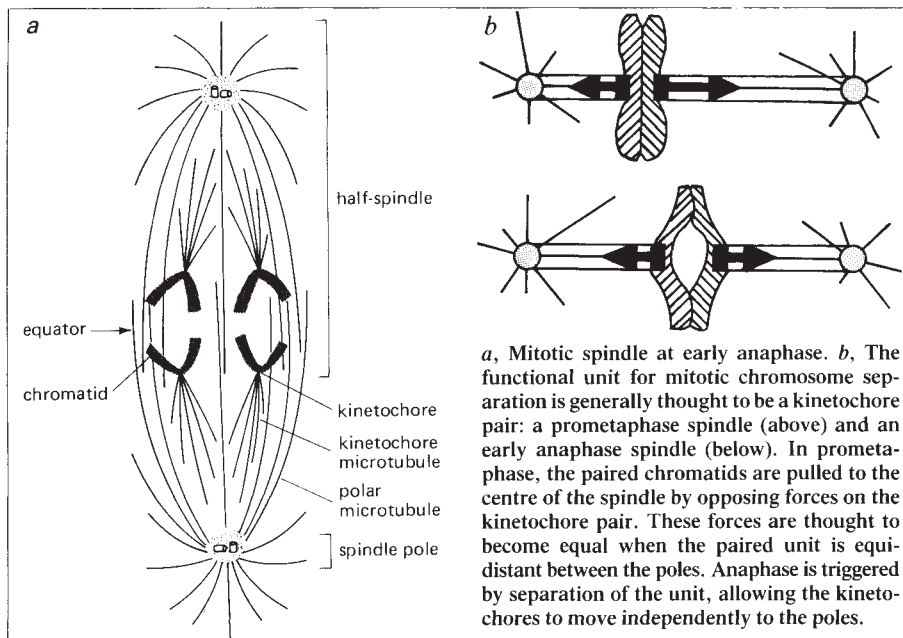
Because it is generally thought that kinetochores choreograph mitotic movements as pairs⁸, one on each chromatid oriented to opposite poles, these movements of free kinetochores present a paradox: either the functional unit in metaphase congression and anaphase segregation is not in fact an opposed pair of kinetochores; or single kinetochores can somehow become pairs in the absence of conventional DNA replication.

The data of Brinkley *et al.* tend to suggest the latter possibility. These authors see an average kinetochore number much higher than that expected for G₁ cells ($2n = 22$). The actual number varies from cell to cell, sometimes exceeding the number expected if every kinetochore had replicated, or otherwise become a pair ($4n = 44$). Furthermore, the efficiency of segregation, that is the tendency of the kinetochores to split into groups of equal size, is highest in cells in which the most kinetochores have split into pairs. This suggests that the unit of congression and segregation is indeed the kinetochore pair, which begs the question of how the duplicated structures arise in the absence of DNA replication. Electron microscopy gives some indication that individual kinetochores can send microtubules towards both poles, and thus congress. Whether such individuals can split and segregate at anaphase is not clear.

How do kinetochore pairs arise in MUGs? One way would be for the kinetochores to duplicate without concomitant centromere DNA replication. Classic experiments in lily by Stern and Hotta⁹ suggested that kinetochore DNA replicates very late, during mitosis, which would mean that the protein structure can duplicate before the DNA. But in yeast the time of centromere replication is in S-phase¹⁰. Whether the results from studying the simple yeast centromeres will turn out to be generally applicable remains to be seen.

Brinkley *et al.* suggest³ instead that the spindle forces fragment the kinetochores in MUGs. Spindle forces are known to be strong¹¹, so fragmentation of an isolated kinetochore may not come as a surprise. This idea should be testable by asking whether the amount of material in the split kinetochore, assayed by quantitative immunofluorescence, is split or duplicated.

What is more of a surprise is that once split, these fragments should remain attached to each other throughout metaphase, and then undergo a seemingly normal separation at anaphase. The nature of the connection which normally holds sister chromatids together before



a, Mitotic spindle at early anaphase. b, The functional unit for mitotic chromosome separation is generally thought to be a kinetochore pair: a prometaphase spindle (above) and an early anaphase spindle (below). In prometaphase, the paired chromatids are pulled to the centre of the spindle by opposing forces on the kinetochore pair. These forces are thought to become equal when the paired unit is equidistant between the poles. Anaphase is triggered by separation of the unit, allowing the kinetochores to move independently to the poles.

anaphase is unknown. The MUG observations suggest that a strong connection which is separable by the anaphase trigger can arise between elements which would never normally be joined in such a manner — that is, a connection can form within a single split kinetochore in a MUG, in a similar fashion to a connection between sister kinetochores in normal cells. This result seems to favour models in which the anaphase-separable connection is made by some distributed feature of chromatin, such as DNA catenation¹², rather than models which require interactions between specific sites on the sister chromatids. □

1. Dietz, R. *Naturwissenschaften* **56**, 237–248 (1969).
2. Mazia, D., Harris, P.J. & Bibring, T.J. *biophys. Biochem. Cytol.* **7**, 1–20 (1960).
3. Brinkley, B.R. *et al. Nature* **336**, 251–254 (1988).
4. Schlegel, R. & Pardoll, A.B. *Science* **232**, 1264–1266 (1986).
5. Murray, A. *Nature* **326**, 542–543 (1987).
6. Brenner, S., Pepper, D., Berns, M.W., Tan, E. & Brinkley, B.R. *J. Cell Biol.* **91**, 95–102 (1981).
7. Earnshaw, W.C. *et al. J. Cell Biol.* **104**, 817–829 (1987).
8. Ostergren, G. *Hereditas* **37**, 85–156 (1951).
9. Stern, H. & Hotta, Y. *Phil. Trans. R. Soc. B* **277**, 277–294 (1977).
10. McCarroll, R.M. & Fangman, W.L. *Cell* **54**, 505 (1988).
11. Nicklas, R.D. & Kubai, D.F. *Chromosoma* **92**, 313 (1985).
12. Murray, A.W. & Szostak, J.W. *A. Rev. Cell Biol.* **1**, 289–315 (1985).

Tim Mitchison and Anthony Hyman are in the Department of Pharmacology, University of California, San Francisco, California 94117, USA.

Geomorphology

Hill slopes and hollows

Mike Kirkby

THE overall scale of most landscapes is set by the density of river channels. The network of these channels is important in controlling the local relief in an area and the height of flood peaks associated with rain storms. Surprisingly, there is no satisfactory theory to explain channel density, particularly for humid areas. On page 232 of this issue¹, however, Montgomery and Dietrich address this question and find a strong inverse dependence of stream-head area on hill-slope gradient, which they relate to a model in which landslides initiate channel growth.

The processes that control stream head area are of fundamental importance for hydrology and geomorphology. Models of catchment hydrological response to storms² increasingly distinguish between hill-slope and channel processes, but how the division arises is not clear. Geomorphologists interpret current landscapes in terms of present and past climates and of topographical changes that result as erosion lowers the landscape.

In valley-head hollows, hydrological and sediment-transport effects interact both on hill slopes and in channels. The hydrological interaction is dynamically dependent on the sequence of major

storms which extend the channels, alternating with quiescent periods during which sediment from the hill slopes tends to refill them. The stream-head and hollow area is thus very sensitive to changes in the climate and land use. Improved understanding could therefore assist environmental reconstructions, over time spans of interest to both archaeologists and Quaternary scientists.

Despite extensive studies^{3–5}, starting with that by Horton⁶, channel density dependencies have not been explained. Density, measured as the total length of channel per unit catchment area, ranges from below 1 km km⁻² in some temperate areas to over 100 km km⁻² in badland areas. In a channel network, about half of the total length is in the unbranched tributaries (Fig. 1). In areas of high channel density, the area of valley hollow needed to sustain each channel head must be small, and conversely, low-density channels drain large areas.

The work by Montgomery and Dietrich, which was carried out in well-vegetated areas in California and Oregon, shows that there is also a strong inverse dependence of head area on slope for streams with sharp head-cuts (Fig. 2). From their results, the authors suggest that small-scale landslides, initiated under predictable conditions of soil saturation and valley slope, cause the channel growth. The model seems to work well for valley axes steeper than 10°–15°.

Other work^{7,8} has indicated similar relationships, by showing that there are empirical thresholds between gullied and ungullied slopes. Melton⁹ had shown that there is a clear link between vegetation cover and the density of drainage channels, at least in the semi-arid zone of the American west. This implies that drainage density falls, and the stream-head area therefore increases in making the transition from sparsely vegetated semi-arid

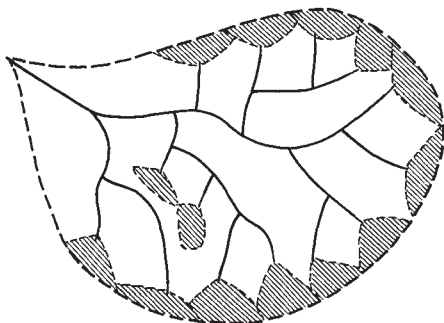


Fig. 1 Schematic map of drainage channels, showing the relationship between channel density, channel spacing and the area drained by each channel head.

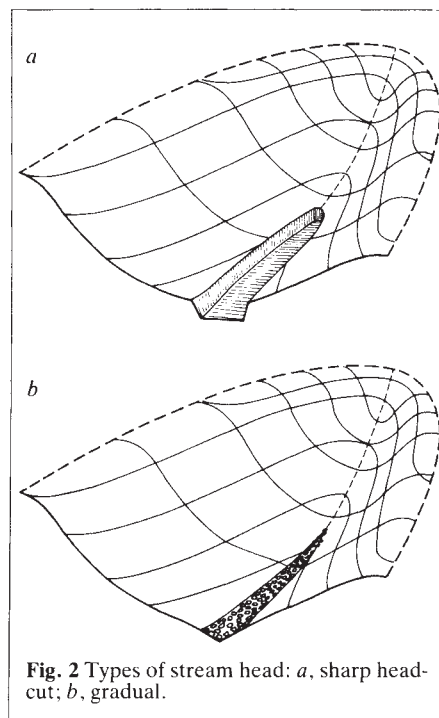


Fig. 2 Types of stream head: a, sharp head-cut; b, gradual.

towards lush humid areas. Montgomery and Dietrich's¹ data imply that, in areas which are slightly more humid than Melton's, stream-head area increases gradually from wetter to dryer areas. Although the effect is weak, it provides some of the first good evidence that channel spacing could be forecast for humid areas.

There is, however, at least one difficulty with these data. The stream heads described all show a sharp head-cut where the channel first forms in the valley floor. In many areas, stream heads are much more gradual, as an undoubted channel gradually merges upwards into a hillside hollow (Fig. 2b). This type of gradual stream head is most common where the material in it consists of coarse gravel or boulders, and is found both in humid areas like the Appalachians and in semi-arid areas like south-east Spain. Not only are gradual channel heads hard to locate precisely, but there is some doubt about the applicability of Montgomery and Dietrich's results to this case, because there is no role for landslides in forming a gradual stream head. Indeed, some preliminary data¹ suggest that stream-head area increases with gradient in some cases. □

1. Montgomery, D.R. & Dietrich, W.E. *Nature* **336**, 232–234 (1988).
2. Hack, J.T. & Goodlett, J.C. *US Geol. Surv. Prof. Pap.* **347** (1960).
3. Melton, M.A. *Office nav. Res. Tech. Rep. 11* Project NR 389–42 (1987).
4. Chorley, R.J. & Morgan, M.A. *Geol. Soc. Am. Bull.* **73**, 17–34 (1962).
5. Schumm, S.A. *Geol. Soc. Am. Bull.* **67**, 597–646 (1956).
6. Horton, R.E. *Geol. Soc. Am. Bull.* **56**, 275–370 (1945).
7. Patton, P.C. & Schumm, S.A. *Geology* **3**, 88–90 (1975).
8. Graf, W.L. *Earth Surface Processes* **4**, 1–14 (1979).
9. Rodriguez-Iturbe, I. & Valdez, J. *Water Resources Res.* **15**, 1409–1420 (1979).

Mike Kirkby is Professor of Physical Geography at Leeds University, Leeds LS2 9JT, UK.

Allosteric enzymes

Control by phosphorylation

M. F. Perutz

PHOSPHORYLATION controls the activity of many enzymes, but muscle phosphorylase is the only one of them whose structure has been solved. It is a dimer of two identical subunits, each containing a single polypeptide chain of 842 amino-acid residues and one pyridoxal phosphate. It is an allosteric enzyme that exists in at least two forms, the active (*a*) and the inactive or less active (*b*) form. Transition from *b* to *a* is brought about by phosphorylation of a single pair of serines (Ser 14). The active form catalyses the reaction of glycogen with inorganic phosphate to form glucose-1-phosphate. Several other metabolites act as allosteric effectors: for example, AMP activates and glucose 6-phosphate (G6P) inhibits catalysis. The structure of phosphorylase *b* coupled to the weak activator IMP has been solved by Louise Johnson's group at the University of Oxford; that of phosphorylase *a* inhibited by glucose by Robert Fletterick's team at the University of California, San Francisco. The two groups have now compared their structures and report their results on page 215 of this issue¹. They show that phosphorylation of Ser 14 induces stereochemical changes at the interface between the two subunits and at the closely spaced binding sites of AMP and G6P, but they cannot yet see how catalysis at the distant pyridoxal phosphate site is activated.

True to Jacques Monod's predictions, enzymes that exhibit cooperative effects and feedback inhibition have been found to be oligomers made up of protein subunits capable of undergoing transitions between two or more structures that differ by the arrangement of the subunits and by the bonds between them. If equivalence is to be conserved in a dimer of identical subunits related by 2-fold symmetry, each subunit can rotate about an axis perpendicular to the 2-fold axis. In addition, the subunits can move in opposite direc-

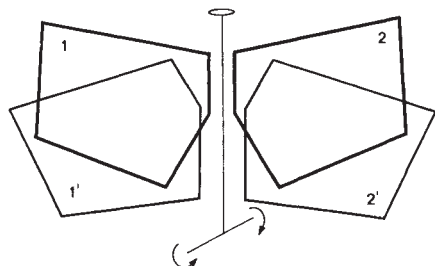


Fig. 1 Possible changes in quaternary structure that preserve equivalence in a dimer of identical subunits related by a 2-fold symmetry axis. I have drawn the rotation axes below the two subunits, to indicate that they can lie anywhere along the 2-fold axis, as long as they are normal to it.

tions parallel to these axes (Fig. 1). The motions observed in phosphorylase conform to these rules, but they are very slight: a rotation by less than 1° is coupled to a translation of only about 0.5 Å.

That change in quaternary structure is surprisingly small in view of the large changes in secondary and tertiary structure observed in parts of the subunit boundary. The most important of these is the burial and ordering of the amino-terminal 16 residues and the exposure and disorder of the carboxy-terminal 5 residues in phosphorylase *a*, and the reversal of these features in phosphorylase *b*. These

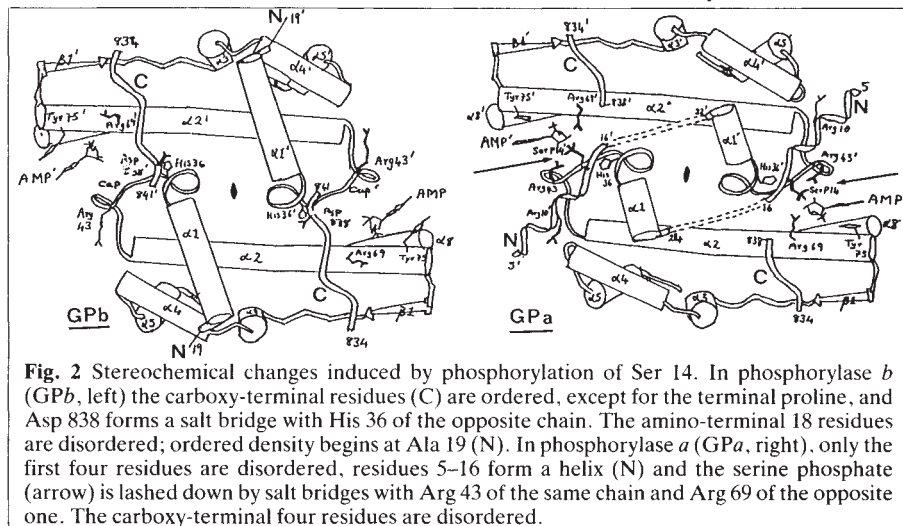


Fig. 2 Stereochemical changes induced by phosphorylation of Ser 14. In phosphorylase *b* (GPb, left) the carboxy-terminal residues (C) are ordered, except for the terminal proline, and Asp 838 forms a salt bridge with His 36 of the opposite chain. The amino-terminal 18 residues are disordered; ordered density begins at Ala 19 (N). In phosphorylase *a* (GPa, right), only the first four residues are disordered, residues 5–16 form a helix (N) and the serine phosphate (arrow) is lashed down by salt bridges with Arg 43 of the same chain and Arg 69 of the opposite chain. The carboxy-terminal four residues are disordered.

movements are accompanied by changes in hydrogen bonding. In *b*, Asp 838 is tied down by a salt bridge to His 36 of the opposite subunit. On transition to *a* that salt bridge is broken, and the histidine rotates about the α - β bond to form a hydrogen bond with a main-chain atom of phosphoserine 14 of the same subunit. The phosphate also forms salt bridges with two arginines, one from the same and the other from the opposite subunit (Fig. 2). Thus the dominant interactions responsible for the allosteric transition are electrostatic. Other changes in salt bridges and non-polar contacts between subunits follow in train, and these are transmitted to the AMP- and G6P-binding site that lies at the subunit boundary only 15 Å from phosphoserine 14. AMP is wedged between an α -helix from one subunit and a non-helical loop from the opposite subunit, referred to as the cap; that loop is separated from phosphoserine 14 by a short helix. Comparison of the two structures shows how dephosphorylation of Ser 14 and transition to phosphorylase *b* weakens both electrostatic and van der Waals interactions with AMP, thus increasing its

dissociation constant from the enzyme 100-fold. The same changes strengthen G6P binding.

I was intrigued to see indications of a hydrogen bond between the amine of an asparagine (Asn 44) and the flat side of the adenine in the complex of AMP with phosphorylase *a* (Fig. 5b of ref. 1 on page 221). This is a previously unrecognized form of hydrogen bond between amines or amides and aromatic rings which I first encountered in a complex of haemoglobin with the drug bezafibrate and which has since been found to make significant contributions to the stability of many protein structures²⁻⁴.

In haemoglobin, which Monod called an honorary enzyme, additional bonds between the subunits constrain the deoxy — or tense — structure and lower its oxygen affinity; this observation led me to believe that the activity of other allosteric

enzymes would be similarly controlled, the inactive form always being the tense and the active one the relaxed structure, but in phosphorylase the reverse seems to be true. Phosphorylase *a* has the greater number of bonds between the subunits, the larger area of buried surface, and is less easily dissociated than *b*. Would this still be true in the absence of inhibition by glucose or does failure of phosphorylase *a* to crystallize without glucose imply that it becomes floppy on activation?

The catalytic sites that contain the pyridoxal phosphates are buried at the ends of 12–15-Å long channels and are more than 30 Å away from the subunit boundary and from the phosphoserine and AMP (Fig. 3). Comparison between the *a* and *b* structures suggests possible pathways of communication between these sites, but has revealed no working mechanism as yet. Each subunit has a 5-turn helix ($\alpha 7$, the tower helix, residues 261–277) extending across the subunit interface. The two tower helices interlock. Each helix is followed by a non-helical loop ending in a reverse turn that contains Asp 283. This aspartate lies close to the

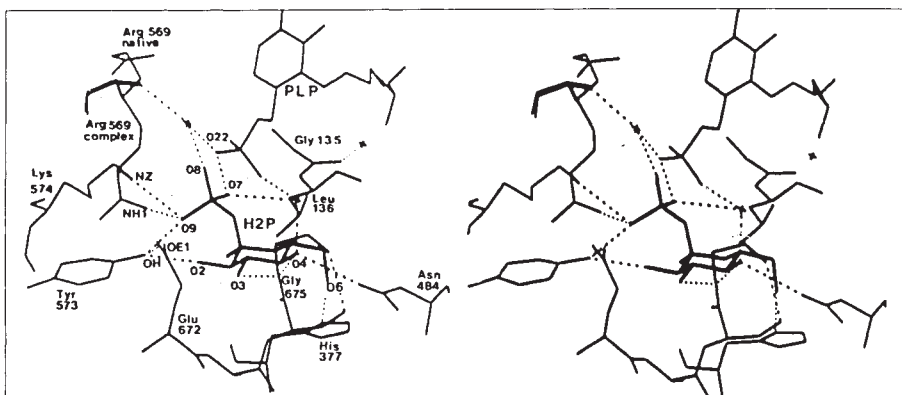


Fig. 4 Stereo diagram showing the active site of phosphorylase *b* with the transition state analogue heptulose-2-phosphate (H2P) and the cofactor PLP bound to Lys 680. Note how Arg 569 has swung into the active site, displacing Asp 283 and Asn 284 (not shown) which block access to the active site in substrate-free phosphorylase *b* (ref. 5).

pyridoxal phosphate and blocks access to it in the inhibited form of phosphorylase *a*. It acts like a lock. If the two tower helices shifted relative to each other on activation, they might pull Asp 283 away from the active site and make it accessible, but no such movement is to be seen as yet. Following the reverse turn, the chain continues as a helix ($\alpha 8$, residues 289–313) which crosses to the AMP-binding site. There is also a helix ($\alpha 2$, residues 47–78) that links the AMP-binding site to the phosphoserine on the opposite sides of the dimer. Thus the structure shows more-or-less direct connections between the distant sites, but does not reveal how stereochemical changes are transmitted along them.

Synchrotron radiation studies of catalysis in crystals of phosphorylase *b* by the Oxford group⁵ indicate some of the changes at the active site likely to accompany transition from the inhibited to the active form of phosphorylase *a*. This group followed the reaction of the sugar heptenitol with inorganic phosphate to form heptulose-2-phosphate in the presence of the weak activator IMP. Heptulose-2-phosphate is a transition-state analogue that remains bound to the active site without undergoing further reactions. The structure of its complex with the enzyme does indeed indicate the expected movement of the 281–288 loop

to make way for the substrate. One of the first chemical findings⁶ one would like to explain is the transition of the pyridoxal phosphate from the di-anion, active form in phosphorylase *a* to the mono-anion, inactive form in phosphorylase *b*. The structures show that the cofactor phosphate interacts with two lysines (568 and 574) in

a and with only one lysine (568) in *b* (Fig. 4). When *b* binds heptulose-2-phosphate, Arg 569 moves from a buried position to another close to the product phosphate, displacing Asp 283. A combination of similar stereochemical changes may also take place on removal of the inhibitor glucose from phosphorylase *a*, and may stabilize the di-anionic phosphate of the cofactor.

Let us hope that the next instalments of this long saga will provide firmer ideas of the catalytic mechanism and also tell us how changes at the active site are triggered by changes at the distant regulator sites. That is the key to the enzyme's allosteric control mechanisms. □

1. Sprang, S. R. *et al.* *Nature* **336**, 215–221 (1988).
2. Perutz, K. F., Fermi, G., Abraham, D. J., Poyart, C. & Bursaux, E. *J. Am. chem. Soc.* **108**, 1064–1078 (1986).
3. Burley, S. K. & Petsko, G. A. *FEBS Lett.* **203**, 139–142 (1986).
4. Levitt, M. & Perutz, M. F. *J. molec. Biol.* **201**, 751–754 (1988).
5. Hajdu, J. *et al.* *EMBO J.* **6**, 539–545 (1987).
6. Helmreich, E. J. M. & Klein, H. W. *Angew. Chemie Int. Ed. Engl.* **19**, 441–445 (1980).

M. F. Perutz is in the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

Fundamental forces

Atoms test electroweak theory

Eugene D. Commins

NOECKER, Masterson and Wieman of the National Bureau of Standards in Boulder, Colorado have just reported¹ the most precise measurement yet of parity non-conservation (PNC) in atomic caesium. This phenomenon arises almost entirely from interference between the neutral weak and electromagnetic interactions that couple a valence electron to the atomic nucleus. The Boulder experiment, as well as others already completed or now being developed, is thus of interest in connection with the standard electroweak model, which describes in a unified way the neutral and charged weak nuclear interactions as well as the electromagnetic interaction.

This model was first formulated by S. Weinberg and A. Salam in 1967. Neutral weak interactions were first observed at CERN in 1973 and subsequent experiments, mainly in high-energy physics but also including atomic PNC, were carried out to elucidate and verify the predictions of the standard model. These culminated in the direct observation by C. Rubbia and colleagues at CERN of the 'carriers' of the charged and neutral weak interactions, the intermediate vector bosons $W^\pm$ and Z^0 , respectively.

Although the standard model is thus firmly established, there are still good reasons to continue atomic PNC experiments, provided that they can be done with sufficient precision in atoms amen-

able to unambiguous theoretical analysis. For one thing, the momentum transferred by the Z^0 between electron and nucleus in an atom is many orders of magnitude smaller than typical momentum transfers in high-energy neutral-weak experiments and it is worthwhile to demonstrate that the standard model holds precisely over a large range in this variable. Furthermore, physicists hope to uncover contributions from terms beyond the dominant one, including effects arising from electroweak radiative corrections and from a nuclear anapole moment, described below.

For these reasons, the new result is noteworthy: it achieves a precision more than a factor of three better than previously obtained with caesium, for which the most precise theoretical atomic calculations are available. Furthermore, the new data reveal the first evidence for the existence of the anapole effect.

To grasp the essential elements of PNC experiments it is convenient to imagine at first that the neutral weak coupling does not exist, and that electron and nucleus communicate only by means of the conventional electromagnetic interaction (photon exchange). This interaction conserves parity — it is mirror symmetric. In a world of atoms governed solely by electromagnetic interaction, any atomic physics experiment and its image viewed in a mirror would give identical results. In the language of quantum mechanics one

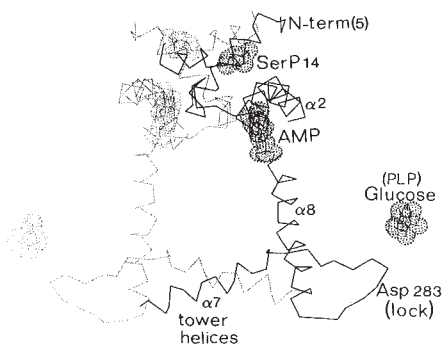


Fig. 3 Structural links between the serine phosphates, the allosteric AMP and the catalytic pyridoxal phosphate (PLP) sites in phosphorylase *a*.

says that the electromagnetic hamiltonian is a true scalar and, as a consequence, each eigenstate of the hamiltonian (for example, each stationary state of the caesium atom, see Fig. 1) is also an eigenstate of parity: the wavefunction either changes sign (is odd) or does not change sign (is even) under mirror inversion. For example, the $6S$ and $7S$ states are even in this approximation and the $6P$ states are odd.

But the much feebleness of neutral weak interaction is not mirror symmetric. In addition to a scalar part, the effective hamiltonian describing it includes a pseudo-scalar portion that changes sign under mirror inversion. Thus the total hamiltonian of the electroweak interaction is not exactly mirror symmetric and any given stationary state is no longer exactly even or exactly odd.

Instead a state of given nominal parity contains a small admixture of states of the opposite parity. For example, the ground state of caesium is predominantly $6S$ but contains in addition a small admixture of $6P_{1/2}$ and also higher $P_{1/2}$ states. The same can be said for the $7S$

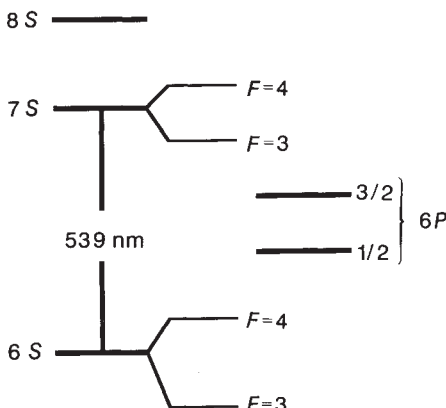


Fig. 1 Low-lying energy levels of caesium (not to scale). The PNC experiment uses the $6S$ – $7S$ transition at 539 nm, the amplitude for which contains a parity-violating E1 component and a Stark-induced E1 component in the presence of an external electric field. Interference between these two components produces the observed effect. A slight difference in the PNC asymmetry between the transitions $F=4$ to $F=3$ and $F=3$ to $F=4$ indicates the presence of a nuclear anapole moment, caused by nuclear parity violation.

state. (A comprehensive analysis of the way in which such admixtures arise from the standard neutral weak coupling was first presented by M.A. Bouchiat and C. Bouchiat^{2–4} of the Laboratoire de Spectroscopie Hertzienne in Paris in 1974; they also proposed many of the most fruitful ideas in this field and M.A. Bouchiat together with co-workers has carried out a meticulous PNC experiment with caesium⁵.)

A transition from the ground $6S$ state to the $7S$ state can be induced by exposing caesium atoms to laser light at 539 nm (Fig. 1). In the approximation where

neutral weak effects are neglected the transition amplitude is of the magnetic-dipole ($M1$) type, not the electric-dipole ($E1$) type, because both initial and final states are even under mirror inversion. Because of neutral weak coupling, however, and the consequence that both states $6S$ and $7S$ contain some admixture of odd states, the transition amplitude contains in addition a small parity-nonconserving electric-dipole component E_p . Furthermore, if the caesium atoms are placed in an external electric field E , the transition amplitude contains an electric-dipole component (βE) by virtue of the Stark effect, quite independent of the neutral weak interaction. In the Boulder experiment, as in the Paris caesium experiment and our own related work on thallium, interference between the amplitudes E_p and βE is measured.

Noecker *et al.*¹ use an elliptically polarized optical standing wave at 539 nm, generated by a laser and a Fabry–Perot interferometer (Fig. 2). The electric field is in the x -direction and the caesium atoms are in an atomic beam that travels in the z -direction. A magnetic field B , also along z , is imposed to split various Zeeman sublevels of the caesium hyperfine levels $F=3,4$ so that the laser can be tuned to individual F, m_F – $F', m_{F'}$ spectral transitions. The desired interference effect causes a transition-rate enhancement for some of these and a suppression for the rest.

The Boulder experiment yields a determination of the ‘mixing angle’ Θ_w , a fundamental parameter in the standard model:

$$\sin^2 \Theta_w = 0.219 \pm 0.007 \pm 0.018$$

where the first uncertainty is experimental and where the second is due to uncertainties in atomic physics calculations. This can be compared with the average of $\sin^2 \Theta_w = 0.230 \pm 0.005$ obtained from diverse high-energy experiments⁶.

Although the bulk of the observed effect arises from neutral-weak–electromagnetic interference as already outlined, the anapole-moment effect contributes to a small difference in the PNC asymmetry for different hyperfine components. This arises from parity violation in the nucleus itself, which causes the nucleus to generate a vector potential of ‘axial vector’ form⁷. Conventional electromagnetic coupling of the valence electron to this axial-vector potential gives rise to the nuclear-spin dependence seen by the Boulder group¹.

In this, as in all other ‘Stark interference’ PNC experiments, good control over systematic errors is maintained, not merely by multiple subtractions and symmetry in the experimental design, but also by auxiliary experiments in which the atoms themselves are used to measure the systematic effects directly. The difficulties arise from misalignment and imperfect reversal of the electric field, misalignment

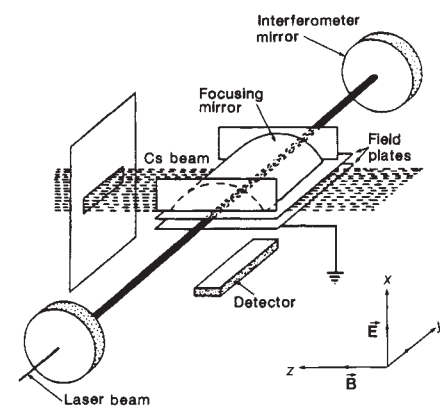


Fig. 2 In the Boulder experiment, a laser beam accurately tuned by a Fabry–Perot interferometer to the $6S$ – $7S$ transition of caesium crosses a beam of caesium atoms in the presence of electric (E) and magnetic (B) fields. The electric field from the field plates induces the Stark effect — a perturbation of atomic levels — which adds an electric-dipole ($E1$) component to the parity-nonconserving $E1$ component of the optical transitions. The transitions are observed by the fluorescence of the excited atoms using the detector.

of the magnetic field, imperfections in the polarization of the incoming photons and from imperfect cancellation of the $M1$ – βE interference^{5,8}.

The Boulder experiment will soon be improved substantially by polarizing the atomic beam before it enters the interaction region. In this way all ground-state atoms are concentrated in one magnetic sublevel (m_F), and a factor of 16 in intensity is gained. Meanwhile, at Paris an ingenious scheme has been proposed⁹ for caesium, in which, in addition to a linearly polarized, pulsed 539-nm laser beam, one also applies pulsed ‘probe’ beams at 1.4 μm . The function of the latter beams is to stimulate emission between the $7S$ and $6P$ states.

An intriguing feature of this novel scheme arises from the fact that a beam of light gains intensity when it traverses a medium with ‘inverted’ populations (having more atoms in the higher, excited state than the lower) and stimulated emission is induced. This opens up the possibility of amplifying the PNC effect. Substantial improvements in other PNC experiments are under way as well, and it is clear that atomic physics still has something interesting and useful to contribute to our knowledge of fundamental interactions. □

1. Noecker, M.C., Masterson, B.P. & Wieman, C.E. *Phys. Rev. Lett.* **61**, 310–313 (1988).
2. Bouchiat, M.A. & Bouchiat, C. *Phys. Lett.* **48B**, 111, (1974).
3. Bouchiat, M.A. & Bouchiat, C. *J. Phys. Paris* **35**, 899 (1974).
4. Bouchiat, M.A. & Bouchiat, C. *J. Phys. Paris* **36**, 493 (1975).
5. Bouchiat, M.A. *et al.* *J. Phys. Paris* **47**, 1709 (1986).
6. Amaldi, U. *et al.* *Phys. Rev. D* **36**, 1385 (1987).
7. Flambaum, V., Khriplovich, I. & Sushkov, O. *Phys. Lett.* **146B**, 367 (1984).
8. Drell, P.S. & Commins, E.D. *Phys. Rev. A* **32**, 2196 (1985).
9. Bouchiat, M.A. *et al.* *Opt. Commun.* **56**, 100 (1985).

Eugene D. Commins is in the Department of Physics, University of California, Berkeley, California 94720, USA.

Plate tectonics

Action and delayed reaction

H.J. Melosh

ON page 234 of this issue¹, Rydelek and Sacks use a long record of earthquakes in north-east Japan to argue that subduction-zone ('sea') earthquakes along the Japan Trench are correlated with earthquakes in the over-riding plate ('land' earthquakes). Although the record's quality is vastly different in the pre- and post-instrumental eras, the apparent correlation is clearly illustrated in a way that will give seismologists much food for thought. The most surprising aspect of the correlation is that the land earthquakes apparently precede sea earthquakes by about 36 years — it has always been assumed that a subduction-zone earthquake occurred first, releasing a large strain pulse that initiated small aftershocks in the adjacent plates.

Rydelek and Sacks's proposal that strain pulses can slowly propagate through the lithosphere is not new. In fact, it is a little-appreciated corollary of plate tectonics that such slow, coherent strain pulses must exist. In its simplest expression, plate tectonics depicts the Earth's crust and upper mantle as comprising a few rigid lithospheric plates that slowly drift relative to one another over a low-viscosity asthenosphere. Interactions between the plates are supposed to occur only at their edges. Within this simplistic view, however, a paradox is concealed, for the model does not explain how the purely local displacements that occur during plate-boundary earthquakes are translated into the global displacements of the plates.

This paradox was resolved in 1967 by Elsasser², who treated the lithospheric plates as elastic objects, hence capable of small deformations. Because the lithosphere is in contact with the viscous asthenosphere, long-lasting strain pulses do not travel at the speed of sound in rock (several kilometres per second), but rather at a much slower rate that depends upon the asthenosphere's viscosity, among other things. Elsasser showed that such strain pulses propagate in a way described by a diffusion equation so that the displacements caused by an individual plate-boundary earthquake slowly spread and blend with the displacements caused by previous and adjacent earthquakes to produce the mean motion of the plates. (This is similar to the way that the diurnal and seasonal temperature changes at the surface of the Earth blend to produce a nearly constant average temperature at depths of a few tens of metres in the soil.) The velocity of each individual strain pulse declines as the inverse square root of time, whereas the combined strain of a

long series of stress pulses propagates at constant speed³. Near the edge of each plate there should thus be a boundary zone several hundred kilometres wide where displacements fluctuate with time and position, reflecting the pattern of the most recent earthquake events, whereas the motion in the interiors of the large plates reflects only the long-term average effect of large numbers of plate-boundary earthquakes.

Previous workers sought the predicted strain-pulse propagation in this boundary zone, looking for evidence of strain or stress changes in the aftermath of large

kilometre mean distance between correlated earthquakes and the 36-year delay is not greatly different from viscosities derived from other strain-propagation observations. The principal innovation in this study is the suggestion that intraplate earthquakes can trigger plate-boundary events. Exactly how this happens is not clear, but Rydelek and Sacks propose that the intraplate earthquake first relieves compressional stress caused by the subducting oceanic plate. When the (relatively) extensional strain pulse from this event arrives at the subduction zone several decades later, the decrease in normal stress across the plate boundary fault increases its proximity to failure and a subduction-zone earthquake can occur.

Because the strain changes that arise in the initial intraplate event are gradual, the actual earthquake is presumably triggered by some more proximal cause: the extensional strain pulse merely enhances the probability of failure. A somewhat surprising aspect of this explanation is that the propagating extensional pulse apparently fails to trigger significant earthquake activity closer to the initiating event. By the time the strain pulse reaches the plate boundary it has not only spread out, but is also much weaker than it was in the vicinity of the first earthquake. How can it propagate 'silently' for four decades before it somehow triggers the plate-boundary event?

Although it is clear that a great deal more work must be done before a full understanding of earthquake interactions is attained, the paper of Rydelek and Sacks may point this work to a class of interactions heretofore neglected. Unfortunately, earthquake correlations do not provide a continuous record of strain-pulse propagation, so that the course of pulse propagation between earthquakes is largely inferential. Furthermore, because Rydelek and Sacks had to exploit more than 400 years of data to demonstrate the correlation between intraplate and plate-boundary earthquakes in Japan, their method may not be applicable to sites where less complete records have been kept. A more comprehensive understanding of the physics of earthquake interactions may have to await dense space and time coverage of postseismic motions using the millimetre-accuracy geodetic techniques now being developed. □

Making waves: long after the damage caused by the 1964 earthquake in Niigata on the west coast of Japan has been repaired, strain pulses continue to propagate slowly through the lithosphere.

earthquakes. Because direct measurements of horizontal displacement changes are not yet available at the required accuracy, the spread of earthquake aftershock patterns into the adjacent plates⁴, secondary earthquakes correlated with major plate-boundary events^{5,6} and changes in post-seismic uplift with time^{7,8} have all been scrutinized for evidence of strain propagation. (But the deployment of high-precision satellite-laser-ranging and very-long-baseline-interferometry systems — see the News and Views article by T. A. Herring⁹ — may soon give direct measurements the necessary accuracy.)

Rydelek and Sacks¹, in showing in this issue that intraplate earthquakes in the over-riding plate precede plate-boundary events, have turned the problem around somewhat, although the basic principle of strain propagation remains the same. The asthenospheric viscosity (7×10^{18} pascal seconds) that they derive from the 200-

Popperfoto

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

1. Rydelek, P.A. & Sacks, I.S. *Nature* **336**, 234–237 (1988).
2. Elsasser, W.M. in *The Application of Modern Physics to the Earth and Planetary Interiors* (ed. Runcorn, S.K.) 223–246 (Wiley, New York, 1967).
3. Bott, M.H.P. & Dean, D.S. *Nature* **243**, 339–341 (1973).
4. Melosh, H.J. *J. geophys. Res.* **81**, 5621–5632 (1976).
5. Anderson, D.L. *Science* **187**, 1077–1079 (1975).
6. Dmowska, R., Rice, J.R., Lovison, L.C. & Josell, D. *J. geophys. Res.* **93**, 7869–7884 (1988).
7. Nur, A. & Mavko, G. *Science* **183**, 204–206 (1974).
8. Thatcher, W. *et al.* *J. geophys. Res.* **85**, 6429–6435 (1980).
9. Herring, T.A. *Nature* **334**, 102–103 (1988).

H. J. Melosh is in the Lunar and Planetary Laboratory and Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA.

C.V. Raman on acoustics

In this, the third of four extracts from Raman's letters on acoustics selected to mark the author's centenary, Raman shows his skill in interpreting complicated experimental data. The figures he presents are of a type now well known as many perform similar studies of the behaviour of metals and plastics under impact. In Raman's day, such work was rare and he shows considerable insight in interpreting his Percussion Figures in *Isotropic Solids*. The reference to the work of Hertz is of interest because this work is usually overshadowed by his studies of electromagnetic phenomena.

J.M. Bowsher, University of Surrey

The accompanying photographs are of interest as illustrating the manner in which an isotropic solid breaks down under the stresses set up by impact when these exceed the limits of perfect recovery, and have a bearing on the theory of the collision of elastic solids developed by Hertz.

Figs 1, 2 and 3 are pictures of the percussion figure, taken from three different points of view, produced on the surface of a thick glass plate by

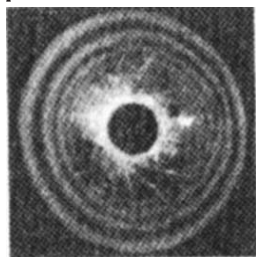


Fig. 1



Fig. 2



Fig. 3

the impact of a polished hard steel ball. Near the centre of the region of contact between the sphere and the plate the stresses are mainly in the nature of a volume-compression, and fracture accordingly does not originate there, but occurs at or near the margin of the compressed area in the form of a fine circular crack which spreads inwards into the plate obliquely in the form of a surface of revolution. This is clearly shown in Fig. 1, which is a front view of the percussion figure by reflected light, the dark circle in the middle being the uninjured area of contact between ball and plate. The circular interference-rings seen in the picture are a measure of the separation of the surfaces of the internal fracture within the plate.

Fig. 2 is a side view, and Fig. 3 an oblique view, of the internal fracture seen through the edge of the plate, the lower half of each picture being the image of the upper half formed by the reflection of light at the interior surface of the plate. The circular area of contact from the margin of which the fracture starts appears in Fig. 3 as an elliptic white disc at the centre. It seems clear that the internal fracture practically occurs along the surface of maximum shearing stress set up during the impact.

From *Nature* 104, 113 (1919).

Ecological genetics

1868 and all that for *Magicicada*

Godfrey M. Hewitt, Richard A. Nichols and Michael G. Ritchie

ON page 237 of this issue¹, Martin and Simon describe the use of mitochondrial DNA markers to unravel the evolutionary origins of periodical cicadas of the genus *Magicicada*. Martin and Simon suggest that an ecological factor caused an extraordinary concerted change from a life cycle of 17 years to one of 13 years throughout a region including much of Illinois and Missouri. If changes in generation time do occur as the authors suggest, their results will help to explain the complex pattern of emergence of these insects and the evolution of their uniquely long life cycles.

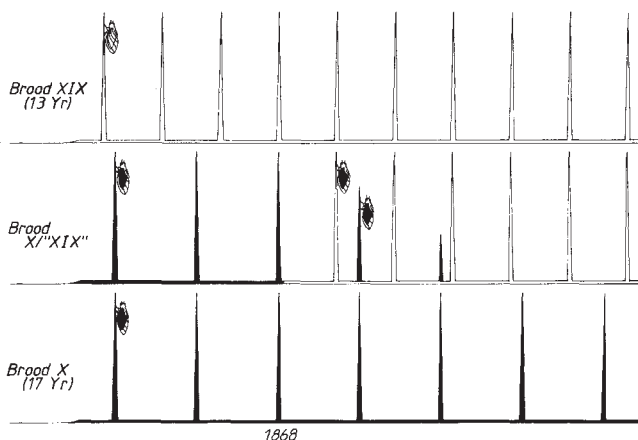
There are three species of *Magicicada* living underground on roots in the deciduous forest region of the eastern United States. In the Mississippi valley and southern states, their life cycle is 13 years, whereas to the north and west it is 17 years. Nearly every year, in some part of the range, adult cicadas emerge synchronously in huge numbers: populations of 1.5 million individuals per acre have been recorded². Cicadas with the same life cycle emerging in the same year are designated a brood. Thirteen of the possible 17 broods of the 17-year cicadas, and three of the 13 possible 13-year broods, have been identified. This synchronized emergence may be advantageous in evolutionary terms because the large numbers satiate predators³, and the long life cycles may prevent predator population sizes becoming synchronized with the emergence time. But if there is strong selection against non-synchronizing individuals, it is difficult to imagine how the life-cycle shift from 17 to 13 years could have happened⁴.

There are records of the emergence of these populations of periodic cicadas going back for more than 100 years. There are reports of small emergences out of step, and evidence that '4-year accelera-

tions' have taken place, in which broods with the same cycle length occur 4 years apart in the same woods. It is interesting that the development of the 17-year cicada involves a 4-year early nymphal period when growth is inhibited; elimination of this period would produce a 13-year life cycle. In Illinois, there is good historical evidence for the disappearance of a 17-year brood and the emergence of a new 13-year brood² (see figure). The emergence dates of these two broods coincide in 1868. Furthermore, the 17-year brood X and the 13-year brood XIX both occur in Illinois, an observation prompting Lloyd *et al.* to suggest⁵ that hybridization between the broods caused the change in the life cycle. The appearance of a small brood in 1898 is consistent with the idea that a single gene controls life cycle: this brood could consist of second-generation (F_2) segregants ($1868 + 13 + 17 = 1898$).

Martin and Simon in this issue¹ show that the 13-year brood in the Illinois area has mitochondrial DNA like that in the adjacent 17-year brood X, and distinct from that of the neighbouring 13-year brood XIX. Furthermore, the new brood is like the 17-year brood in abdominal colour and frequency of the enzyme phosphoglucose mutase (PGM). (This enzyme is one of three polymorphic enzymes analysed by Martin and Simon to provide information on phylogenetic relationships.) Consequently, Martin and Simon argue for an environmentally induced switch in life cycle from 17 to 13 years, and they suggest that high population density may have triggered this switch.

The observed changes in life cycle could therefore result from either hybridization involving dominance for a 13-year cycle, as Lloyd and collaborators suggest, or from a density-induced 4-year acceleration, as suggested by Martin and Simon. There are problems with both



Brood XIX (13-year cycle) to the south emerged together with brood X (17-year cycle) from the north in 1868, and in the region of overlap a 13-year brood replaced the 17-year brood. Dark individuals represent broods with colouring, PGM frequency and mitochondrial DNA typical of brood X.

suggestions, for example, why has the cycle not reverted to 17 years as it appears to have done in other areas?

Evaluation of other evolutionary explanations for the phenomenon depends on more knowledge of distribution, density and dispersal. The apparent absence of nuclear hybrids and brood X in the switched region could result from underground larval competition between different broods so that only one survives. Such competition could interact with predation on the adults, which would eliminate all but the most numerous brood. Alternatively, if a reorganized genome emerged in the F_2 generation at a time of population contraction, it could expand in subsequent generations to fill the region it now occupies⁶. Such dramatic changes in population density and distribution appear plausible: the historical record seems to show marked range expansion and contractions⁴.

The long generation times of these cicadas limit experimentation. It would be interesting to know if the 13- and 17-year broods can mate, and whether they are interfertile. Opportunities for attempting to cross 13- and 17-year broods are rare. Apparently the next window is 1998, when both brood IV(17) and brood XIX(13) emerge; in 2024, broods XIII(17) and XIX(13) emerge, and broods X(17) and XIX(13) not again until 2089. This is indeed an example of a project that requires long-term funding. □

1. Martin, A.P. & Simon, C. *Nature* **336**, 237–239 (1988).
2. Dybas, H.S. & Davis, D.D. *Ecology* **42**, 432–444 (1962).
3. Lloyd, M. & Dybas, H.S. *Evolution* **20**, 133–149 (1966).
4. Lloyd, M. & Dybas, H.S. *Evolution* **20**, 466–505 (1966).
5. Lloyd, M., Kritsky, G. & Simon, C. *Evolution* **37**, 1162–1180 (1983).
6. Hewitt, G.M. *Trends ecol. Evol.* **3**, 158–167 (1988).

Godfrey M. Hewitt, Richard A. Nichols and Michael G. Ritchie are in the School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Alzheimer's disease

Arresting memory decline

T.W. Robbins

H. G. WELLS once wrote a short story entitled "The New Accelerator" which describes the potential consequences of introducing into society drugs that enhance intellectual prowess. The discovery of specific patterns of neurochemical depletion in the brains of patients dying of Alzheimer's disease and other neurodegenerative dementias now places a special, but perhaps no less fantastic, premium on the discovery of drugs that can ameliorate or arrest declining cognitive function. Promising and sobering prospects emerged at a recent meeting* held to assess the state of this enterprise.

A well-known drug, nicotine, was reported to enhance certain aspects of test performance in patients early in the course of Alzheimer's disease, whereas two novel compounds arising from drug-development programmes with strikingly different strategies are found to enhance aspects of cognitive function in the laboratory, but have yet to be tested clinically. What also became clear is that positive results in animal tests of memory are far more frequent than reliable effects in humans, and that extrapolations from laboratory tests of cognition in normal human volunteers to the clinic should be made with caution.

Alzheimer's disease has been largely resistant to treatment. In keeping with the cholinergic hypothesis of the cognitive deficits in dementia — that memory decline is related to loss of cholinergic

neurons in the brain — the most efficacious drugs have been anticholinesterases such as physostigmine and tetrahydroaminoacridine, drugs that prevent the breakdown of the neurotransmitter acetylcholine. The effects of these drugs in controlled studies have often been positive but are too small to be of general clinical significance. Plausible reasons for the lack of convincing effects range from problems with the drugs (toxicity, side-effects and poor pharmacokinetics) to the infeasibility of restoring function by enhancing one neurotransmitter substance in patients with several forms of neuropathology, and heterogeneity in the cognitive deficits themselves.

An alternative logical pharmacological strategy has been to develop cholinergic receptor agonists. There are two main kinds of acetylcholine receptors, muscarinic and nicotinic, both of which are located at pre- and post-synaptic sites. However, the strategy of developing more specific muscarinic receptor agonists has not, so far, been crowned with any great degree of success, as confirmed at the meeting by R. Spiegel's (Sandoz Ltd, Basel) review of clinical experience with the muscarinic agonist RS-86.

Another, more novel approach focused on nicotine receptors. Large reductions in nicotine receptor binding have been measured in the cortex and hippocampus of patients dying of Alzheimer's disease, so a trial with nicotine would thus seem worth attempting. B. Sahakian (Institute of Psychiatry, London) described the preliminary results of a laboratory study of

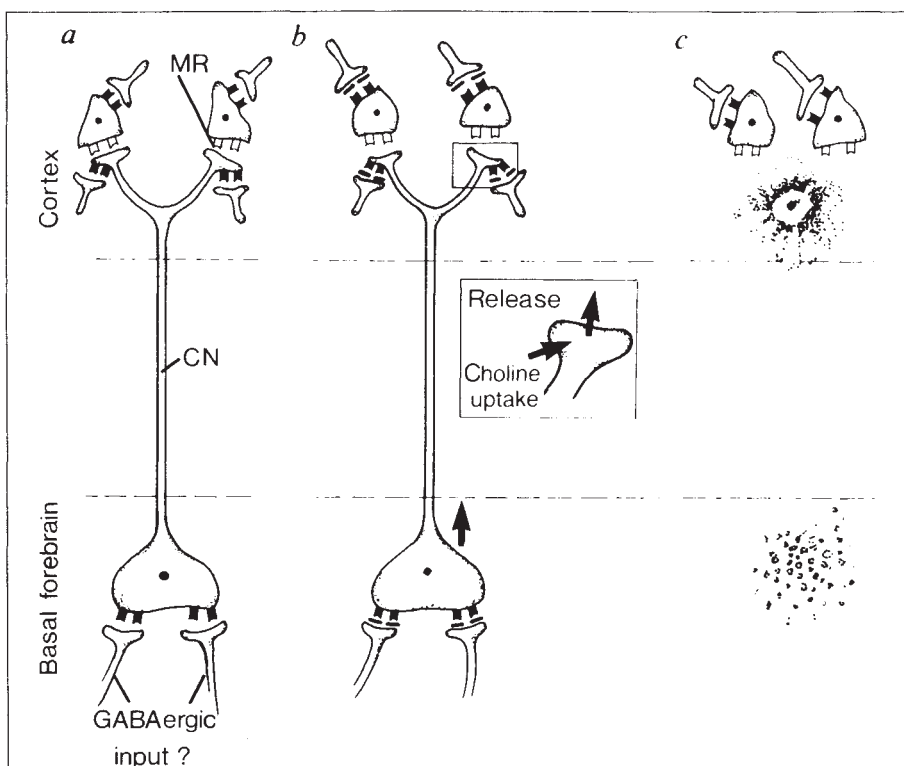
the effects of acute subcutaneous administration of a range of low doses of nicotine. Three groups of subjects were tested: normal young adults; normal elderly subjects; and seven cases of subjects early in the course of probable dementia of the Alzheimer type (DAT), as diagnosed by published criteria. Tests of rapid visual information processing, involving working memory and attention, as well as visual short-term memory based on an animal paradigm, and an index of cortical arousal (critical flicker fusion or CFF), were used. The results indicate significant dose-related improvements in performance in young volunteers and in the DAT subjects in the test of rapid information processing and CFF, but no improvement in the test of visual short-term memory. These experiments are now being repeated using a larger sample, and the use of nicotine chewing gum is being planned in a chronic study that will include some assessment of activities of everyday living.

The promising results with nicotine raise questions about underlying mechanisms. The pattern of effects suggests general improvements in attention and arousal that could be related to a mild stimulant action. It is possible that some of the behavioural effects of nicotine in the rat are mediated via the mesolimbic dopaminergic system, and thus would resemble the effects of conventional psychomotor stimulants such as methylphenidate and amphetamine. But there are side-effects of psychomotor stimulants which preclude their use in treatment of dementia. Another possibility is that nicotine acts at the presynaptic receptors of surviving cholinergic terminals to facilitate release of acetylcholine.

A second strategy for developing cognitive enhancers also depends upon boosting presynaptic cholinergic function, but in a rather indirect way. There is now considerable evidence that benzodiazepines impair cognitive function, producing, for example, symptoms of anterograde amnesia, presumably via their action at the benzodiazepine/GABA receptor complex. M. Sarter (Schering AG, Berlin) described a rationale based on identifying β -carboline compounds acting as benzodiazepine receptor partial inverse agonists or antagonists that could be expected to have the opposite effects. Partial inverse agonists (such as FG 7142) or antagonists (such as ZK 93426) could have beneficial effects on cognition, while avoiding the undesirable convulsant and anxiogenic effects of the full inverse agonists.

Sarter reported evidence that such compounds could act at receptors in the substantia innominata, in the region of the magnocellular cholinergic neurons of the nucleus basalis, and thus indirectly enhance cortical cholinergic function by inhibiting an inhibitory GABAergic

* Cognitive Enhancers: Preclinical and Clinical Pharmacology Satellite symposium of the second EBPS Meeting, Athens, 29 August 1988.



The β -carbolines may offer a treatment for senile dementia primarily by disinhibiting the remaining cholinergic neurons of the basal forebrain. *a*, In the normal brain, cholinergic neurons (CN) may innervate predominantly cortical pyramidal cells. *b*, Possible therapeutic action of the antagonist β -carbolines ZK 93426 in facilitating the action of cholinergic neurons in the early stages of Alzheimer's disease. All benzodiazepine receptors are blocked (-), and acetylcholine release and high-affinity choline uptake are enhanced. Such a disinhibitory effect at the cell body of cholinergic neurons enhances impulse conduction, and also possibly the transport of proteins and phospholipids. *c*, By the time of death, the majority (30–70 per cent) of the extrinsic cholinergic neurons have degenerated without a major consequence for the postsynaptic muscarinic receptors (MR). It is not known whether cholinergic therapy would retard this degenerative process. (Adapted from Sarter, M. *et al. Trends Neurosci.* 11, 13; 1988.)

influence on the cholinergic cells (see figure). This hypothesis is supported by the fact that ZK 93426 reverses some of the deleterious behavioural effects in rats of the anticholinergic drug scopolamine and the effects of neurotoxic lesions of the substantia innominata (now, however, widely suspected to produce many of their behavioural effects by disrupting non-cholinergic mechanisms).

ZK 93426 has not been clinically assessed, but its effect on attention and memory has been tested in normal young human volunteers, with acute intravenous administration. T. Duka (Schering AG, Berlin) reported some significant effects on measures of vigilance and attention, as well as small beneficial effects on visual recognition memory. No significant effects were found on tests of rapid information processing similar to those used to test nicotine above. It seems that the β -carboline strategy might be worth pursuing, but a more potent compound would be required to hold much promise for patients with DAT.

A radically different approach, with a strong flavour of serendipity, emerged from the report by G. Clincke (Janssen Research Laboratories, Beerse) of fairly impressive effects on sophisticated tests of

learning and memory, in normal young and elderly volunteers, of a compound whose pharmacological actions are quite obscure. The drug concerned, R58735, has a profile related to the anti-epileptic agent carbamazepine, and has some affinity for histamine (H₂) receptors, but the relevance of these actions to its effects on cognition is dubious.

Although effective cognitive enhancers may yet emerge from rational programmes of drug development based on what we know about the neurochemical changes in dementia and a careful characterization of drug action at a cellular level, this putative class of compounds may be no less prone to unexpected additions than the progenitors of the antipsychotic and antidepressant classes that were discovered almost accidentally in the 1950s. In the meantime, as potential but discerning consumers, we should perhaps be prepared for the anticipated inundation of candidate cognitive enhancers by developing standardized and comprehensive tests of cognition that can be used for comparing their effects across species as well as in objective clinical assessment. □

T.W. Robbins is in the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

Daedalus

Private lights

LAST week Daedalus devised a highly efficient lighting system for TV studios. Each camera projects light from its lens on the studio scene, which scatters it to the cube-corner-patterned studio walls. These reflect each photon straight back to the point on the scene from which it was originally scattered. It then undergoes a second scattering, the precise optical reverse of the first, and is therefore fired back into the camera lens. Almost all the light from the camera lamps returns to make the image, and mere watts of light can create a brilliant and colour-saturated scene. Daedalus is now extending the idea to human vision.

His lamp-bearing beam-splitter spectacles launch light out from the wearer's eyes. A few sheets of corner-cube-embossed reflector film placed around the scene then appear to illuminate it brilliantly in vivid colours. As a way of saving domestic electricity this is perhaps a bit cumbersome, but for lighting public spaces it is perfect. The old wasteful socialist ideal of public lighting would be replaced by the modern efficient entrepreneurial concept of private lighting. For the intriguing feature of Daedalus's system is that each user sees only by his own light. Photons from other observers bounce back to them alone, and hardly interfere. People who like their illumination blindingly bright, or garishly coloured, or polarized, or strobed, can all have it, without annoying anyone else. Those too poor to afford the equipment are, very properly, left in the dark. Only the reflectors are public property, and they waste no energy.

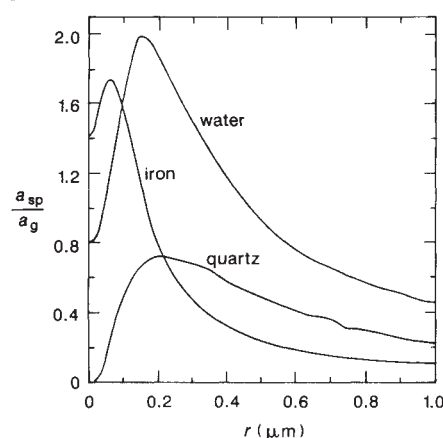
The new system should be particularly useful for lighting the roads. Daedalus is devising a cube-corner-printing steam-roller which embosses the desired reflecting pattern on the carriageway, and aluminizes it for high reflectivity. Wearers of his beam-splitter spectacles will then see the nocturnal road as a shining ribbon of light on which other vehicles are sharply outlined, illuminated from below. Intriguingly, this system may be usefully speed-sensitive. A photon bouncing off a moving vehicle will find it displaced by a few nanometres when it returns from the reflective road. Its second reflection will not be quite the reverse of its first one, and it may not quite retrace its path to the eye. The resulting strange and non-intuitive pattern of light and shade will immediately call attention to the moving vehicle. At first Daedalus feared that a very fast vehicle might be rendered completely invisible by this effect. But he now realizes that it will still obscure light from the road behind it, and will appear as an ominous speeding shadow.

David Jones

Microorganism transport from Earth to Mars

SIR—In a recent letter, Melosh¹ presented the concept that large impacts on the Earth may eject rocks containing microorganisms that could fall on Mars or other planets. I have independently developed a similar model, and in November 1987 I submitted a very long article, which *Nature* was unable to publish. Here I would like to compare the two models. I will use the term dust/solar pressure/keplerian model (DSPK) to refer to my work and boulder/keplerian model (BK) when referring to Melosh's work; the meanings of these terms will become evident in what follows.

The DSPK model focuses on the transport of microorganisms from the Earth to Mars. This model starts with a meteorite impact on the Earth that can launch into space debris containing microorganisms from the surface of the planet. A fraction of this ejecta goes into orbit around the planet and some small fraction reaches



Ratio of sunlight-pressure acceleration (a_{sp}) to solar gravitational attraction (a_g) as a function of particle radius (r) for various materials⁴.

escape velocity. The small particles ($<0.5 \mu\text{m}$) in the fraction that escapes the gravitational influence of the planet and is placed in Earth's orbit beyond $45R_{\oplus}$ (where $R_{\oplus}$ is the Earth radius) can reach Mars by means of solar light pressure as a propulsion mechanism²; larger particles may reach Mars by means of keplerian trajectories. The converse of the latter can be seen in the case of the SNC meteorites, which are thought to have come from Mars.

The timescale for the transport of microscopic particles from Earth to Mars by the solar light pressure mechanism is about two months, whereas the boulder-sized rocks proposed by Melosh as the carriers of microorganisms take, on average, a few million years to reach Mars. Laboratory and other evidence do not support the idea that microorganisms could last in viable form for a million years in such a hostile environment. Furthermore we can expect that radiation from radioactive elements present in the rock

itself would destroy the microorganisms over such long periods of time. The mechanism of propulsion by solar radiation pressure is much more efficient as a transport mechanism, but places the constraint that only particles with a diameter of $0.5 \mu\text{m}$ or less will be transported⁴ (see figure). Since there are many microorganisms with diameters of $\sim 0.01 \mu\text{m}$, each dust particle could transport up to 50 such entities.

Another important feature considered in the DSPK model is that particles with diameters between $0.5 \mu\text{m}$ and 1 cm that have reached escape velocity are many orders of magnitude more abundant than boulder-sized rocks and are distributed over a dust cloud several thousand kilometres in diameter when leaving the Earth. The probability of impact of such a cloud on Mars is up to $\sim 10^{10}$ times greater than for the boulders of $\sim 1 \text{ m}$ in diameter proposed by Melosh. Another important difference between the two models is that the DSPK model includes a consideration of the environment met by the transported microorganisms. The presence of liquid water on Mars in the past⁵ suggests that at least at some time the transported microorganisms may have met with an environment favourable to their survival. Future Mars missions may thus find either fossils or live microorganisms and other forms of life in the north polar region, where there are still substantial amounts of water ice.

MIGUEL A. MORENO

Uriburu 26 Sur,
Pocito 5427, San Juan,
Argentina

1. Melosh, H.J. *Nature* **332**, 687–688 (1988).
2. Belton, M.J.S. *Science* **151**, 35–44 (1966).
3. Mitchell, F.J. & Ellis, W.L. *Analysis of Surveyor 3 Material and Photographs Returned by Apollo 12* 239–248 (1972).
4. Shapiro, I.L., Lautman, D.A. & Colombo, G. *J. geophys. Res.* **71**, 5695–5704 (1966).
5. Carr, M. *Icarus* **68**, 187–216 (1986).

Did God make RNA?

SIR—The recent dispute between Maizels and Weiner¹ and Benner and Ellington² on the properties of primeval 'ribo-organisms' illustrates the extent to which the picture of a protein-less 'RNA world' has captured the imagination of molecular biologists. What was only a cautiously advanced hypothesis some years ago has now gained so much credibility as to go largely unquestioned. It is even presented to students as "almost certain"³. Some measure of balance needs to be restored, at least for the benefit of those who are not experts in the field.

If RNA came first, how did it arise? This crucial question is rarely raised. When it is, the astonishing admission is made that there is no plausible answer. Even committed proponents of the RNA-

first creed acknowledge that "the accumulation of substantial quantities of relatively pure mononucleotides on the primitive Earth is highly implausible"⁴. As a way out, the authors postulate the prior formation of a somewhat less demanding oligonucleotide analogue which, however, still poses a formidable challenge to prebiotic ingenuity. Until the RNA world is provided with more solid foundations, its present vogue hardly seems justified. At the very least, appropriate consideration deserves to be given to alternative theories, of which there are several. A brief outline of some proposals that I formulated recently^{5–8} is given below.

My key argument is that nothing short of an extensive collection of catalysts, similar, be it in very crude form, to present-day enzymes, could have accounted for the prebiotic synthesis of such complex molecules as nucleotides. Given this premise, my model requires as sole starting materials the kind of simple amino acids and other small organic molecules that are readily obtained in abiotic simulation experiments or found in meteorites. Further chemical development is all accomplished through catalysis, leading to the progressive emergence of a network of protometabolic pathways, which I visualize as not very different from existing metabolism. Thus, the model, unlike most others, makes minimal demands on undirected abiotic chemistry.

I see the catalysts that are the main props of the model as rare members of a large population of randomly assembled oligopeptides and other multimers, from which active molecules could have emerged selectively thanks to substrate stabilization⁵. I view them as the primitive protoenzymes that must, for statistical reasons, be postulated as the precursors of present-day enzymes (see, for example, ref. 9). They could be ancestral to 'mini-proteins' now encoded by exons (see, for example, ref. 10). I imagine their assembly as having proceeded from thioesters, which energetic considerations indicate could have been present in small, but significant, amounts in the kind of sulphur-rich, hot and acidic medium that many authors identify as the probable cradle of life (see, for example, ref. 11). Such thioesters could also have fuelled all the endergonic steps of protometabolism.

The possibility that a 'thioester world' may have preceded the RNA world is consistent with biochemical knowledge, as well as with the views of many biochemists. Indeed, the thioester bond lies at the very heart of life. It is the precursor of the pyrophosphate bond in key substrate-level phosphorylation processes, of virtually all naturally occurring ester bonds, and of numerous carbon-carbon bonds, such as citrate, fatty acids porphyrins, terpenes and steroids. It is particularly

suggestive that the thioester bond is also the precursor of peptide bonds in a number of bacterial products that happen to be heterogeneously constructed from D as well as L α -amino acids, and sometimes other precursors, as must have been the multimers postulated by my model. Thus, the thioester bond, born in hot, acidic sulphurous waters, could have been the central prebiotic source of both catalysis and energy, allowing the spreading of a long and complex chain of metabolic events that eventually led to something that could be called an RNA world, except that it would have been richly endowed with peptides.

CHRISTIAN DE DUVE

International Institute of Cellular and Molecular Pathology,
75.50 avenue Hippocrate 75,
B-1200 Brussels, Belgium

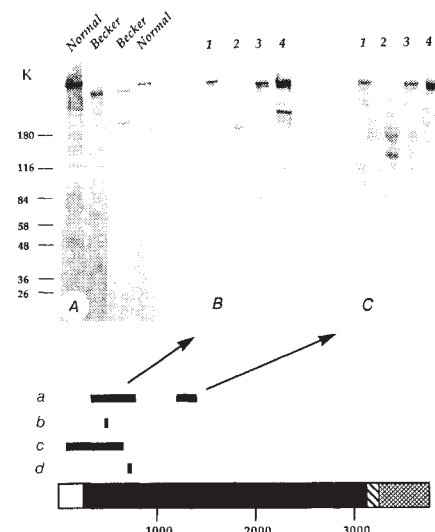
1. Maizels, N. & Weiner, A. M. *Nature* **330**, 616 (1987).
2. Benner, S. A. & Ellington, A. D. *Nature* **332**, 688 (1988).
3. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. & Weiner, A. N. *Molecular Biology of the Gene*, 4th edn, Vol. II, 1104–1163 (Benjamin/Cummings, Menlo Park, 1988).
4. Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4398–4402 (1987).
5. De Duve, C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8253 (1987).
6. De Duve, C. in *The Roots of Modern Biochemistry* (eds Kleinkauf, H., von Döhren, H. & Jaenicke, L.) 881–894 (Gruyter, Berlin, 1988).
7. Du Duve, C. *Nature* **333**, 117–118 (1988).
8. De Duve, C. *Blueprint for a Cell* (Patterson, Burlington, in the press).
9. Ninio, J. *Approches Moléculaires de l'Evolution* (Masson, Paris, 1979).
10. Gilbert, W. *Cold Spring Harbor Symp. quant. Biol.* **52**, 901–905 (1987).
11. Woese, C. R. *Microbiol. Rev.* **51**, 221–271 (1987).

Proteolytic fragment or new gene product?

SIR—A recent report by Arahata *et al.*¹ has apparently created some confusion regarding the protein product of the Duchenne muscular dystrophy (DMD) gene. In an attempt to clear up this confusion, we would like to discuss some recent published data, and offer alternative explanations for the results presented by Arahata *et al.* which are more consistent with these data.

First, the DMD messenger RNA contains coding information for a protein of relative molecular mass (M_r) of 427,000 (427K)², a size in complete agreement with that of our previously reported product of the DMD gene, 'dystrophin'.³ Contrary to this, Arahata *et al.* report a relative molecular mass of the DMD protein as 210K, without offering an explanation for how such a protein becomes half the predicted size.

Second, three independent laboratories have used peptide sequences derived from the same or similar messenger RNA sequences as those used by Arahata *et al.* to produce antisera which recognize the predicted 427K protein, called 'dystrophin' (see accompanying figure)^{3,5}. Third, there is a large body of data which strongly correlates abnormalities of dystrophin (427K)



Bottom, a schematic drawing of the 427K dystrophin protein with its four distinct domains³. Above this are the regions of the dystrophin protein to which polyclonal antisera have been produced. (a, ref. 3; b, ref. 1; c, ref. 4; d, ref. 5). All antisera recognize a common 427K protein (dystrophin), except that characterized by Arahata, *et al.*¹. Top, A shows two normal human muscle samples and two samples obtained from patients with Becker muscular dystrophy. Dystrophin (427K) is seen to be of abnormal (smaller) relative molecular mass in both Becker patients. The second patient exhibits a 200K degradation product of dystrophin which is characteristic of approximately 50% of Becker patients. B and C are duplicate immunoblots incubated with anti-dystrophin (427K) antisera directed against the indicated peptides³. Lane 1, normal human total muscle protein; lane 2, total muscle protein from a patient diagnosed as spinal muscular atrophy type III; lane 3, a second, fresh, biopsy obtained from this same patient; lane 4, homogenization and crude fractionation of the muscle biopsy shown in lane 3. The samples shown in lanes 2 and 4 exhibit non-specific degradation of dystrophin which resemble the immunoblot pattern of Arahata *et al.* There is also a slightly smaller smooth-muscle-specific isoform of dystrophin migrating immediately below the predominant striated muscle isoform⁹. Marker sizes are derived from pre-stained proteins (Sigma).

with Duchenne and Becker muscular dystrophies at both the genetic and biochemical levels. Perhaps the most convincing of these are the dystrophin abnormalities associated with the milder 'Becker' alleles of DMD. Most Becker patients are predicted to possess 'in frame' exon deletions of the DMD gene which would result in a mRNA encoding a smaller DMD protein⁶. Indeed, the majority of Becker muscular dystrophy patients have a dystrophin of lower M_r in their muscle (ref. 7 and accompanying figure). Such results are impossible to explain unless dystrophin (427K) is the principal product of the DMD gene.

The immunofluorescent data provided by Arahata *et al.* for the 210K protein are in complete agreement with those obtained for dystrophin (427K) by two independent

laboratories^{4,8}. It is therefore likely that their antisera does detect dystrophin. However, the immunoblot results presented closely resemble those we have obtained when protein samples have been exposed to partial proteolysis (B and C of figure). Lane 2 contains total muscle protein from a previously frozen diagnostic muscle biopsy, and exhibits apparently 'abnormal' dystrophin of either 200K (B) or 200K and 150K (C), depending on the antibodies used. Because such an abnormal dystrophin pattern was unique in more than 200 patient biopsies that we have tested, a second, fresh, muscle biopsy was obtained from the same patient and similarly tested for dystrophin (lane 3). This new biopsy contained completely normal dystrophin. We conclude that the original biopsy had undergone partial, non-specific, proteolysis of dystrophin.

We believe that the extensive fractionation protocol employed by Arahata *et al.* resulted in a similar degradation of dystrophin. Although we find it impossible to replicate the fractionation conditions used by this group with the experimental detail given, we think it is significant that the homogenization of the fresh human biopsy (B and C, lane 3) followed by crude fractionation in the presence of protease inhibitors results in the appearance of proteolytic fragments of dystrophin of approximately 200K (lane 4). Such proteolytic fragments are of similar size to those reported by Arahata *et al.* They interpret the differing biochemical subcellular fractionation characteristics of the 210K protein as evidence for its non-homology with dystrophin, but it is difficult to equate the fractionation characteristics of a degradation product with that of its parent protein. We have previously reported that many Becker patients also exhibit a degradation product of dystrophin, similar in size to that shown by Arahata *et al.* (ref. 7, second Becker patient in accompanying figure.) We therefore suggest that the 210K protein is a proteolytic fragment of 427K 'dystrophin'.

ERIC P. HOFFMAN

LOUIS M. KUNKEL

Children's Hospital,
Harvard Medical School and
Howard Hughes Medical Institute,
Boston, Massachusetts 02115, USA

ROBERT H. BROWN JR
Massachusetts General Hospital,
Boston, Massachusetts 02114, USA

1. Arahata, K. *et al.* *Nature* **333**, 861–863 (1988).
2. Koenig, M. *et al.* *Cell* **53**, 219–228 (1988).
3. Hoffman, E. P. *et al.* *Cell* **51**, 919–928 (1987).
4. Zubrzycka-Gaarn, E. E. *et al.* *Nature* **333**, 466–469 (1988).
5. Kao, L., Kristenankys, J., Mendell, J., Rammohan, K. W. & Gruenstein, E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4491 (1988).
6. Monaco, A. P., Bertelson, C. J., Liechti-Gallati, S., Moser, H., Kunkel, L. M. *Genomics* **2**, 90–95 (1988).
7. Hoffman, E. P. *et al.* *New Engl. J. Med.* **318**, 1363 (1988).
8. Bonilla, E. *et al.* *Cell* **54**, 447–452 (1988).
9. Hoffman, E. P. *et al.* *Neuron* **1**, 411–420.

Superconductivity near 70 K in a new family of layered copper oxides

R. J. Cava, B. Batlogg, J. J. Krajewski, L. W. Rupp, L. F. Schneemeyer, T. Siegrist, R. B. vanDover, P. Marsh, W. F. Peck, Jr, P. K. Gallagher, S. H. Glarum, J. H. Marshall, R. C. Farrow, J. V. Waszczak, R. Hull & P. Trevor

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

A new family of high-temperature superconductors is described, with the general formula $Pb_2Sr_2ACu_3O_{8+\delta}$. Although they have the planes of CuO_5 square pyramids characteristic of the other copper-oxide superconductors, the new compounds belong to a distinct structural series, with wide scope for elemental substitution. Their unusual electronic configuration also gives new insight into the role of charge distribution among the structural building blocks in controlling superconductivity.

SINCE the first observation¹ of high-transition-temperature (high- T_c) superconductivity in La-Ba-Cu-O, progress in the understanding of this remarkable phenomenon has been coupled to the discovery of new materials. Until now, three families of copper-oxide-based high- T_c superconductors have been identified, based on $(La,M)_2CuO_4$, $LnBa_2Cu_3O_7$, and $(Ti,Bi)_m(Ba,Sr)_2Ca_{n+1}Cu_nO_{m+2n+2}$ (ref. 2). (Here M represents a metal cation that may substitute on some La sites, and Ln represents a lanthanide.) Here we report the discovery of a new family of planar copper-oxide superconductors with general formula $Pb_2Sr_2ACu_3O_{8+\delta}$ (where A is a lanthanide or a mixture of Ln+Sr or Ca), and describe the synthesis, crystal structure and properties of prototype compounds. We find, for example, that one preliminary optimal composition $Pb_2Sr_2Y_{0.5}Ca_{0.5}Cu_3O_8$ has a superconducting T_c of 68 K. The new family displays the same kind of rich substitutional chemistry as is observed for $LnBa_2Cu_3O_7$, with the phase forming for Y and at least La, Pr, Nd, Sm, Eu, Gd, Dy, Ho, Tm, Yb and Lu, spanning the entire rare-earth series. Wide ranges of large-metal-atom solid solution and oxygen stoichiometry are observed, suggesting many possible avenues to be explored for the optimization of superconducting properties.

Superconductivity is induced in the host compounds $Pb_2Sr_2LnCu_3O_{8+\delta}$ ($\delta=0$) either by partial substitution of a divalent ion (such as Sr or Ca) on the lanthanide site, or possibly by the accommodation of excess oxygen ($\delta>0$), or a combination of both. The compounds can be synthesized only under mildly reducing conditions, which are necessary to maintain Pb in a 2+ oxidation state. Oxidation of $\delta=0$ compounds is possible, but only at low temperatures, where decomposition to a Pb(IV)-containing perovskite is sluggish. Remarkably, the formal average oxidation state of copper in the superconductors is less than 2+, but a clear structural distinction between different types of copper layers leads us to hypothesize that holes are nonetheless present on electronically active CuO pyramidal planes.

Synthesis

The preparative conditions for the new materials are considerably more stringent than for the previously known copper-based superconductors. Direct synthesis of members of this family by reaction of the component metal oxides or carbonates in air or oxygen at temperatures below 900 °C is not possible because of the stability of the oxidized $SrPbO_3$ -based perovskite. Successful synthesis is accomplished by the reaction of PbO with pre-reacted (Sr, Ca, Ln) oxide precursors. The precursors are prepared from oxides and carbonates in the appropriate metal ratios, calcined for 16 hours (in dense Al_2O_3 crucibles) at 920–980 °C in air with one intermediate grinding. Some of the

$Pb_2Sr_2LnCu_3O_{8+\delta}$ compounds can be prepared in air from $PbO + LnSr_2Cu_3O_x$ precursor mixtures, which are not reacted at temperatures below ~850 °C. For example, single-phase $Pb_2Sr_2YCu_3O_{8+\delta}$ ($\delta=0$) can be prepared by reacting PbO with $YSr_2Cu_3O_x$ at 920 °C for 1 h, followed by quenching. Slower cooling results in partial decomposition through oxidation. Short reaction times are generally sufficient to obtain single-phase products. The same air-heating/quenching process does not appear to work, however, for $Pb_2Sr_2LaCu_3O_{8+\delta}$ or $Pb_2Sr_2LuCu_3O_{8+\delta}$.

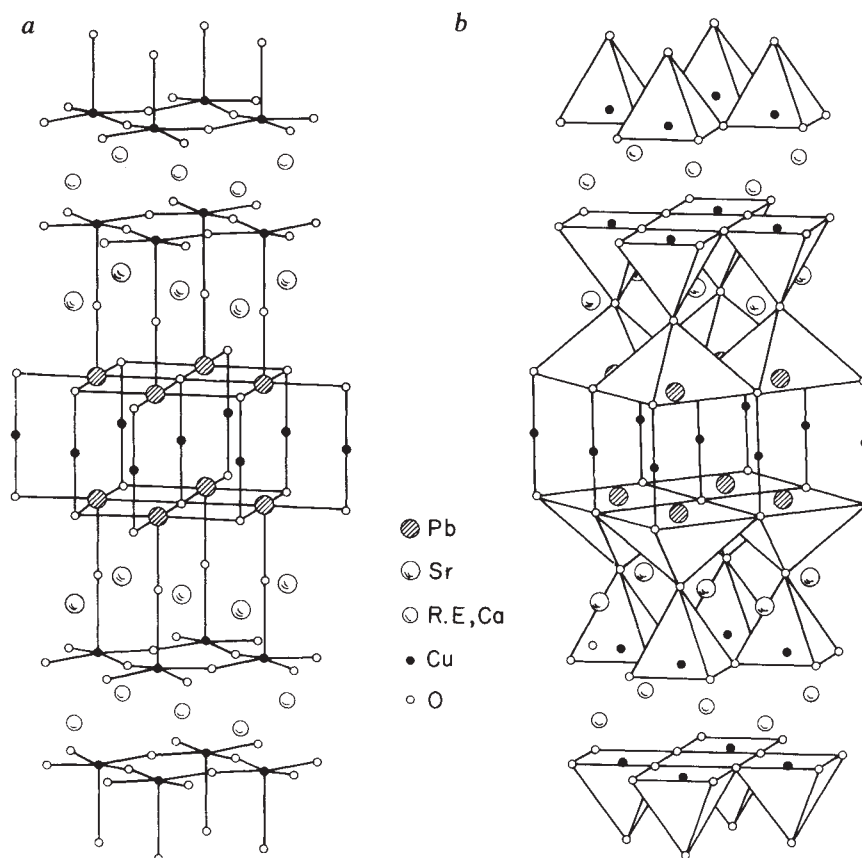
The best synthetic conditions found so far involve the reaction of PbO with the cuprate precursors in thoroughly mixed pressed pellets. Reaction temperatures are between 860 and 925 °C, for times between 1–16 h, in a flowing gas stream of 1% O_2 in N_2 , a mildly reducing atmosphere. For $Pb_2Sr_2Y_{1-x}Ca_xCu_3O_{8+\delta}$, for example, single-phase materials are obtained for $0 \leq x < 0.5$ in 1% O_2 after heating overnight at 865 °C and cooling in the gas stream to room temperature in 15 minutes. Using higher temperatures, higher p_{O_2} in the gas stream or higher Ca contents of the starting mixture results in the intergrowth of 123-type $YSr_2(Pb,Cu)_3O_x$ with the new compound, or the formation of an $SrPbO_3$ -based second phase. Similar procedures are successful for other Sr/rare-earth/Ca combinations. The oxygen contents of $Pb_2Sr_2Y_{1-x}Ca_xCu_3O_{8+\delta}$ for $0 \leq x \leq 0.50$, prepared under these conditions, are measured by reduction in H_2 and are uniformly $\delta=0 \pm 0.1$. Ca is employed as a dopant on the Ln site because it has an ionic size similar to the intermediate rare-earths. We have not yet found synthetic conditions under which $Pb_2Sr_{2+x}Ln_{1-x}Cu_3O_{8+\delta}$ solid solutions can be prepared as single-phase polycrystalline samples that are good bulk superconductors, although superconducting single crystals of that stoichiometry have been prepared.

Single crystals of the superconducting compounds were grown from PbO- and CuO-rich melts using a similar precursor technique. Melt compositions were generally $Pb_{3.5}Sr_3YCu_4O_x$. Following a 30-min soak at 1,025 °C, samples were cooled at 2 °C min^{-1} in the 1% O_2 atmosphere to temperatures between 800 and 400 °C, and were then rapidly cooled to room temperature in the same gas stream. Crystals are plate-like in habit, but are generally more equiaxed than those of $LnBa_2Cu_3O_7$.

Stoichiometry and crystal structure

Compounds of stoichiometry $Pb_2Sr_2LnCu_3O_8$ ($\delta=0$) are not bulk superconductors, although we often observe small amounts of superconductivity (1% or less) in materials of that stoichiometry prepared either by the quench or by the 1%- O_2 synthetic techniques. The non-bulk superconductivity may be due to inhomogeneities in either oxygen content or Sr/Ln distribution.

Fig. 1 Two representations of the crystal structure for the new superconducting compounds, for the case of $\text{Pb}_2\text{Sr}_{2.24}\text{Nd}_{0.76}\text{Cu}_3\text{O}_{8+\delta}$. Representation *a* emphasizes the Cu-O and Pb-O bonding scheme, and representation *b* emphasizes the manner in which Cu-O and Pb-O coordination polyhedra are arranged.



The range of oxygen contents possible for these compounds is remarkable. $\text{Pb}_2\text{Sr}_2\text{YCu}_3\text{O}_{8+\delta}$, $\delta=0$, for example, can be oxidized by heating in O_2 to temperatures below 500°C for short times (2–4 h) to δ values of ~ 1.6 , retaining the same basic crystal structure. We have observed values as large as $\delta=1.8$ for $\text{Pb}_2\text{Sr}_2\text{Y}_{0.75}\text{Ca}_{0.25}\text{Cu}_3\text{O}_{8+\delta}$. Oxidation at temperatures higher than 500°C , or for longer reaction periods, generally results in decomposition to the SrPbO_3 -based perovskite. Powder samples of $\text{Pb}_2\text{Sr}_2\text{YCu}_3\text{O}_{8+\delta}$ with large values of δ are not superconducting. Single crystals of the $\text{Pb}_2\text{Sr}_{2+x}\text{Ln}_x\text{Cu}_3\text{O}_{8+\delta}$ type are superconductors with transition temperatures between 10 and 70 K. These crystals may have non-zero values of δ but have not yet been fully characterized. The range of T_c s observed suggests a complex and interesting relationship between T_c , δ and the $\text{Sr}:\text{Ln}$ ratio.

Powder X-ray diffraction indicates that the new phases have an orthorhombic unit cell which is based on a many-layer perovskite structure. The characteristic X-ray pattern for the prototype compound $\text{Pb}_2\text{Sr}_2\text{YCu}_3\text{O}_8$ is presented in Table 1. The compound deviates only slightly from tetragonal symmetry. The simplest cell consistent with the X-ray pattern is *c*-centred, with lattice parameters $a=5.40$, $b=5.43$, and $c=15.74$ Å. Systematic absences are consistent with a *c*-centred cell down to the detectability limit of 1% maximum intensity. The orthorhombic cell gives an excellent fit to the powder diffraction pattern but a hint of a shoulder on the high 2θ side of the 314 reflection indicates that the true symmetry may be weakly monoclinic. Although the lattice parameters for this family of compounds are very similar to those reported for $\text{TlBa}_2\text{Ca}_2\text{Cu}_3\text{O}_8$ (ref. 3), the crystal structures are quite different. Electron microscope investigations indicate that for some crystals, weak (but sharp) reflections are present which violate the *c*-centring. Furthermore, these studies show the presence of long-period, long-range-ordered superlattices in the *a*-*b* plane, suggesting that a variety of structural distortions and stoichiometry-driven atom-ordering schemes can occur.

The crystal structure of compounds in this family, determined for a superconducting Nd-based single crystal of approximate stoichiometry $\text{Pb}_2\text{Sr}_{2.24}\text{Nd}_{0.76}\text{Cu}_3\text{O}_{8+\delta}$ (determined by structure refinement) is shown in Fig. 1. The crystal employed in the structural determination was twinned, as expected from the pseudo-tetragonal symmetry. The atomic coordinates are reported in the *c*-centred orthorhombic cell to be consistent with the powder data, but a primitive cell with *a* and *b* rotated by 45° and reduced by $\sqrt{2}$ gives an equally good description of the single-crystal data. The very small scattering cross-section of oxygen precludes determination of δ by refinement. The data are well fitted by the structural model (refinement parameter $R=3.7\%$), but a microscopic explanation of the orthorhombic symmetry is not apparent; if the origin is primarily in the oxygen sublattice we would not be able to detect it in the X-ray structure determination.

The basis of the structure comprises infinite planes of corner-shared CuO_5 pyramids separated by eight-coordinate rare-earth atoms, as are common to all the presently known copper-based superconductors with $T_c > 50$ K. The four in-plane copper-oxygen distances are ~ 1.9 Å, and the distance to the apical oxygen is ~ 2.3 Å, both of which are very similar to those observed in $\text{YBa}_2\text{Cu}_3\text{O}_7$. The structural components unique to the new class of materials are the $\text{PbO}-\text{CuO}_5-\text{PbO}$ planes shown in the centre of the Fig. 1. For $\delta=0$, Pb has a distorted flattened square pyramid coordination (sharing edges with adjacent pyramids), with the lone pair pointing toward the vacant sixth site of the coordination octahedron. The PbO_5 pyramids are separated by a single copper layer, which, for $\delta=0$, is oxygen-free, and displays an O-Cu-O coordination characteristic of Cu^{1+} (Cu-O distance ~ 1.8 Å), as is observed in non-superconducting $\text{YBa}_2\text{Cu}_3\text{O}_6$. During the low-temperature oxidation process, oxygen is apparently accommodated in this copper layer, resulting in a large expansion of the *c* axis. The PbO_5 and CuO_5 pyramidal planes are joined by the common oxygens at their apices. The Sr atoms are coordinated to nine oxygens, as in (La,

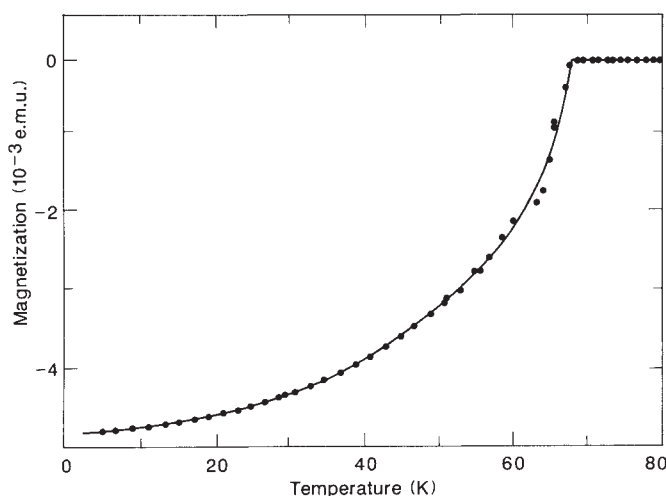


Fig. 2 Magnetization data (d.c. field-cooled at 25 Oe) for $\text{Pb}_2\text{Sr}_2\text{Y}_{0.5}\text{Ca}_{0.5}\text{Cu}_3\text{O}_8$.

$\text{Sr})\text{CuO}_4$, and the Ln site is eight-coordinate, as in the $\text{LnBa}_2\text{Cu}_3\text{O}_7$ family, sandwiched between the CuO_5 pyramidal planes. In the superconducting compound $\text{Pb}_2\text{Sr}_2\text{Y}_{1-x}\text{Ca}_x\text{Cu}_3\text{O}_{8+\delta}$, Ca partially substitutes for Y in the eight-coordinate site.

The crystal structures of all the known copper-oxide-based superconductors are generally described as many-layered perovskites. The similarities and differences among them are most easily illustrated in terms of the stacking sequences of rocksalt-like (AO) and perovskite-like (BO_2) layers². Taking, for example, representatives from the superconductor families that have double CuO_5 pyramidal layers, the stacking sequences are:

$\text{Pb}_2\text{Sr}_2(\text{Y}, \text{Ca})\text{Cu}_3\text{O}_{8+\delta}$
 $-(\text{Y}, \text{Ca})-\text{CuO}_2-\text{SrO}-\text{PbO}-\text{CuO}_8-\text{PbO}-\text{SrO}-\text{CuO}_2-(\text{Y}, \text{Ca})-$

$\text{Ti}_2\text{Ba}_2\text{CaCu}_2\text{O}_8$
 $-\text{Ca}-\text{CuO}_2-\text{BaO}-\text{TlO}-\text{TlO}-\text{BaO}-\text{CuO}_2-\text{Ca}-$

$\text{YBa}_2\text{Cu}_3\text{O}_{6+\delta}$
 $-\text{Y}-\text{CuO}_2-\text{BaO}-\text{CuO}_8-\text{BaO}-\text{CuO}_2-\text{Y}-$

Table 1 Characteristic X-ray powder diffraction pattern for $\text{Pb}_2\text{Sr}_2\text{YCu}_3\text{O}_8$

<i>hkl</i>	<i>d</i>	<i>I/I</i> ₀	<i>hkl</i>	<i>d</i>	<i>I/I</i> ₀
001	15.74	7	116	2.164	11
002	7.87	3	025	2.057	12
003	5.25	2	205	2.050	10
004	3.94	10	008	1.967	7
110	3.831	11	220	1.915	25
111	3.722	24	118, 009	1.750	2
112	3.444	1	027, 207	1.730	1
005	3.148	11	224	1.722	2
113	3.094	11	130	1.717	2
114	2.745	100	310, 131	1.708	3
020	2.717	43	311	1.699	2
200	2.701	43	225	1.636	3
021	2.677	7	133	1.632	3
201	2.662	7	313	1.625	1
006	2.623	6	028	1.593	11
023	2.412	1	208, 119	1.591	11
203	2.401	1	134	1.574	18
024	2.236	2	314	1.568	14
204	2.227	1			

Cu $K\alpha$ radiation, 0–60° 2 θ *c*-centred orthorhombic cell, preliminary indexing, true symmetry may be weakly monoclinic. Lattice parameters $a = 5.4019(15)$, $b = 5.4333(15)$, $c = 15.7388(33)$.

The new superconductors, then, can be seen to be intimately related in structure to those previously described. They can be considered as related to $\text{Ti}_2\text{Ba}_2\text{CaCu}_3\text{O}_8$ by insertion of a single CuO_8 layer between adjacent polarizable AO layers, or related to $\text{YBa}_2\text{Cu}_3\text{O}_{6+\delta}$ by sandwiching of the CuO_8 'chain' layer by two PbO layers. We believe that it is the electronic screening of the CuO_2 planes from the CuO_8 layers by the PbO layers that makes the new superconductors of considerable interest. Furthermore, we expect these materials to be even more anisotropic in their physical properties than those previously known, as the double pyramidal CuO_2 -A- CuO_2 layers are widely separated.

Superconducting properties

We have studied the composition dependence of the superconducting properties of compounds in the series $\text{Pb}_2\text{Sr}_2\text{Y}_{1-x}\text{Ca}_x\text{Cu}_3\text{O}_8$ for $0 \leq x \leq 0.75$, by estimating the flux expulsion measured on cooling in a field of 25 Oe in a d.c. SQUID magnetometer (S.H.E. model 905). The greatest flux expulsion occurs for $x = 0.5$, and is ~20% of the ideal value (see Fig. 2). Because flux becomes trapped in the pores of these low-density ceramics, this is an underestimate of the true volume fraction of superconductivity. For $x \geq 0.5$, the materials were not entirely single-phase, with one or more impurity peaks having a maximum intensity of 5% of the strongest peak in the powder X-ray pattern. This, coupled with the estimate of the volume fraction of superconductivity, suggests that the optimal superconducting composition may have x somewhat greater than 0.5. This could be achieved if different synthetic methods can be found that allow a larger range of solid solution to be attained. We have measured the normal-state susceptibility (in a 20-kOe field) for temperatures below 400 K of apparently single-phase samples (no unindexed X-ray lines to 0.5% maximum intensity) of the non-superconducting endmember $\text{Pb}_2\text{Sr}_2\text{YCu}_3\text{O}_8$ and superconducting $\text{Pb}_2\text{Sr}_2\text{Y}_{0.625}\text{Ca}_{0.375}\text{Cu}_3\text{O}_8$. The susceptibility of the superconductor (χ) is essentially temperature independent ($\chi \approx 1 \times 10^{-4}$ e.m.u. per mole formula unit), with only a slight decrease at low temperatures. This temperature dependence is similar to that of high-quality $\text{YBa}_2\text{Cu}_3\text{O}_7$, and is characterized by the absence of a Curie-Weiss contribution. Furthermore, this supports our conclusion that the copper atoms between the PbO layers are Cu^{1+} . Post-oxidation at 500 °C results in oxidation of this copper to magnetic Cu^{2+} . $\text{Pb}_2\text{Sr}_2\text{YCu}_3\text{O}_8$ appears to be magnetic ($-0.5 \mu\text{B}$ per Cu atom), but further studies are necessary to clarify whether this is intrinsic or is due to the presence of highly magnetic impurity phases that are undetectable by X-ray diffraction.

Figure 3 shows the temperature dependence of the resistivity for a single crystal of $\text{Pb}_2\text{Sr}_2\text{Dy}_{1-x}\text{Ca}_x\text{Cu}_3\text{O}_{8+\delta}$. The midpoint of the superconducting transition is at 51.5 K (indicated by an arrow in Fig. 3), although there is a small foot which gives a zero-resistance T_c of 46 K. Above T_c the temperature

Table 2 Crystallographic data for $\text{Pb}_2\text{Sr}_{2.24}\text{Nd}_{0.76}\text{Cu}_3\text{O}_{8+\delta}$

Atom	Position	<i>x</i>	<i>y</i>	<i>z</i>	<i>B</i> _{iso} [Å ²]
Pb	4l	1/2	0	0.38858 (4)	1.09 (2)
Sr	4k	0	0	0.22184 (9)	0.74 (4)
Nd, Sr*	2a	0	0	0	0.69 (3)
Cu1	2d	0	0	1/2	0.86 (9)
Cu2	4l	1/2	0	0.11074 (13)	0.46 (5)
O1	4l	1/2	0	0.2546 (8)	1.5 (5)
O2	4k	0	0	0.384 (3)	13 (5)
O3	8m	1/4	1/4	0.0995 (5)	0.9 (3)

Orthorhombic cell (pseudotetragonal substructure); $a = 5.435(1)\text{Å}$, $b = 5.463(1)\text{Å}$, $c = 15.817(3)\text{Å}$; space group *Cmmm*, $z = 2$; observed reflections 707, $R_w = 0.037$.

* Mixed occupancy site: 0.24(1) Sr, 0.76(1) Nd.

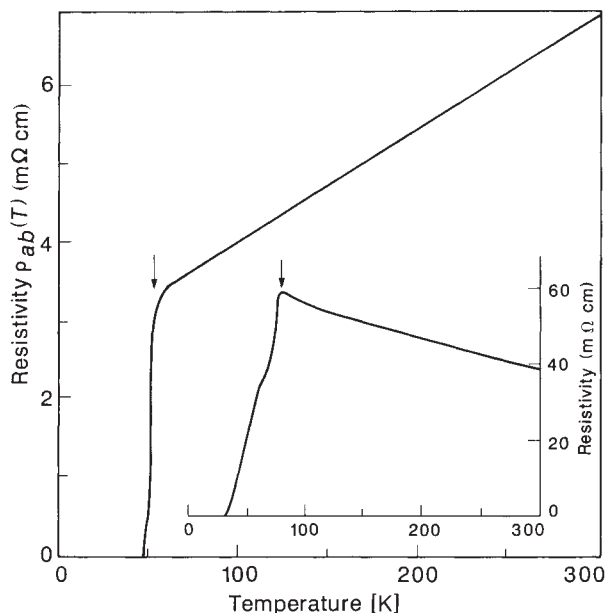


Fig. 3 Resistivity in the a - b plane as a function of temperature for a single crystal of $\text{Pb}_2\text{Sr}_2(\text{Dy,Ca})\text{Cu}_3\text{O}_{8+\delta}$. Inset, typical temperature-dependent resistivity for a polycrystalline sample of $\text{Pb}_2\text{Sr}_2(\text{Y,Ca})\text{Cu}_3\text{O}_8$.

dependence is fairly linear, but near T_c there is a region of positive curvature which, along with the resistivity foot, we attribute to small inhomogeneities in the metal and/or oxygen distribution. The scale of the resistivity is a factor of ten greater than for previous oxide superconductors. It is not yet clear whether this is an intrinsic property.

A typical resistivity curve for a ceramics sample is shown in the inset to Fig. 3, illustrating the typically broad transitions observed. The transition in this sample begins at 79 K (arrow) but zero resistance is achieved (within instrumental accuracy) as 32 K. Note that the resistivity scale is again quite high. We attribute the breadth of the transition and the negative normal-state temperature coefficient to inhomogeneity in the metal and/or oxygen distribution, rather than to exogenous phases at the grain boundaries. The behaviour of this system seems to be very similar to that of $(\text{La,Sr})_2\text{CuO}_4$ (ref. 4).

Electronic aspects

Given that the average formal copper valence of previously known superconductors has always been greater than +2, the new superconductors are unique and, at first sight, anomalous. For the series $\text{Pb}_2\text{Sr}_2\text{Y}_{1-x}\text{Ca}_x\text{Cu}_3\text{O}_8$, the average formal copper valence increases from 1.67 in the non-superconducting $x=0$ member to ~ 1.92 at the maximum Ca concentration studied. At our current estimate of the optimal superconducting composition ($x=0.5$), the average formal valence is 1.83. The linear coordination of the copper atom sandwiched between the PbO sheets, characteristic of Cu^{1+} , and the probable electronic isolation of this layer from the conducting CuO pyramidal planes, imply that the formal charge formulation becomes $\text{Pb}_2\text{Sr}_2\text{YCu}^{1+}\text{Cu}_2^{2+}\text{O}_8$ in the non-superconducting compound. When Ca is substituted for Y, we propose that holes are accommodated only in the CuO_5 planes, and at the $x=0.5$ stoichiometry the formal charge formulation becomes $\text{Pb}_2\text{Sr}_2\text{Y}_{0.5}\text{Ca}_{0.5}\text{Cu}^{1+}\text{Cu}_2^{2.25+}\text{O}_8$, which is consistent with the cur-

rent assumption for previously known high- T_c materials that holes are present in the CuO_5 pyramidal planes.

For $\text{Pb}_2\text{Sr}_2\text{ACu}_3\text{O}_{8+\delta}$ compounds with $\delta > 0$, excess oxygen must be accommodated near the Cu^{1+} planes, and a more complex hole-doping scheme may be operating. We expect that in that case the compound does not respond in a simple fashion to the change in charge through doping of a rigid band; the oxygen inserted in the bonding neighbourhood of the reduced Cu and Pb ions may create the electronic states in which the charge is partly or fully accommodated.

This new family of compounds has a unique crystal structure, yet it also reflects a concept common to all copper-oxide-based superconductors. By now it is well established that superconductivity is associated with layers of Cu-O octahedra, pyramids and squares. The remaining structural building blocks are seen as the electron acceptors which induce the holes necessary for superconductivity in the Cu-O layers. For $\text{YBa}_2\text{Cu}_3\text{O}_{6+\delta}$, for example, we have shown in detail how the CuO_8 chains act as charge reservoirs, and how superconductivity depends on charge transfer between chains and planes⁵.

To illustrate the concept of local charge distribution, one may rewrite the formulae of the high- T_c copper-oxide superconductors as follows: $\text{YBa}_2\text{Cu}_3\text{O}_6[\text{CuO}_8]$; $\text{Sr}_2\text{CaCu}_2\text{O}_6[\text{Bi}_2\text{O}_2]$; $\text{Ba}_2\text{CaCu}_2\text{O}_6[\text{Ti}_2\text{O}_2]$; $\text{Sr}_2(\text{Y,Ca})\text{Cu}_2\text{O}_6[\text{Pb}_2\text{CuO}_{2+\delta}]$; where the structural components in square brackets act as reservoirs which control the charge on the superconducting Cu-O planes. The $\text{PbO-CuO}_8\text{-PbO}$ reservoir layer is likely to be exceptionally flexible in accommodation of charge, and we therefore expect that a relationship between T_c and oxygen stoichiometry as unusual as that for $\text{YBa}_2\text{Cu}_3\text{O}_{6+\delta}$ will eventually be observed. The wide ranges of metal-atom and oxygen-atom stoichiometries in this new family of superconductors are of considerable interest, and warrant further study with the aim of understanding and optimizing the superconducting properties.

We thank D. W. Murphy and K. Rabe for helpful discussions.

Received 21 October; accepted 28 October 1988.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189-193 (1986).
2. Santoro, A., Beech, F., Marezio, M. & Cava, R. J. *Physica C* (in the press).
3. Parkin, S. S. P. *et al. Phys. Rev. Lett.* **61**, 750-753 (1988).

4. Van Dover, R. B., Cava, R. J., Batlogg, B. & Rietman, E. A. *Phys. Rev.* **B35**, 5337-5339 (1987).
5. Cava, R. J. *et al. Physica C* (in the press).

Structural changes in glycogen phosphorylase induced by phosphorylation

S. R. Sprang*, K. R. Acharya†, E. J. Goldsmith*, D. I. Stuart†, K. Varvill†, R. J. Fletterick‡, N. B. Madsen§ & L. N. Johnson†||

* Department of Biochemistry and University of Texas Southwestern Medical Centre, Dallas, Texas 75235-9050, USA

† Laboratory of Molecular Biophysics, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

‡ Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA

§ Department of Biochemistry, University of Alberta, Edmonton, Canada T6G 2H7

A comparison of the refined crystal structures of dimeric glycogen phosphorylase b and a reveals structural changes that represent the first step in the activation of the enzyme. On phosphorylation of serine-14, the N-terminus of each subunit assumes an ordered helical conformation and binds to the surface of the dimer. The consequent structural changes at the N- and C-terminal regions lead to strengthened interactions between subunits and alter the binding sites for allosteric effectors and substrates.

PROTEIN phosphorylation is the most widely-occurring reversible post-translational modification used in the regulation of biological processes^{1,2}. These processes include glycogen metabolism and a number of other metabolic pathways, neuronal mechanisms, hormonal responses, viral oncogene activities and muscle contraction. Protein kinases and phosphatases are controlled by extracellular hormones and by intracellular second messengers (for example cyclic nucleotides, Ca^{2+} and diacylglycerol); phosphorylation provides a mechanism by which intracellular events can be modulated by extracellular signals. It is generally assumed that protein phosphorylation stabilizes different conformational states of the regulated and regulatory molecules to enhance or inhibit biological activity. To date, there has been no direct evidence about the magnitude or characteristics of these conformational changes. Here, we describe a conformational transition induced by phosphorylation of an allosterically regulated enzyme, glycogen phosphorylase (GP), the first phospho-regulated system to be elucidated³.

Glycogen phosphorylase catalyses the degradative phosphorylation of glycogen to glucose-1-phosphate (G1P), the initial step in the generation of metabolic energy in muscle and free glucose in liver. The equilibrium between several conformational states ranging from R (catalytically active; high affinity) to T (inactive; low affinity) is modulated by effectors and phosphorylation⁴⁻⁹. The dephosphorylated enzyme, phosphorylase b (GPb) is virtually inactive *in vivo* but can be activated by AMP or IMP. GPb exhibits both intrasubunit and heterotropic interactions for binding of substrates and AMP. The main physiological activation is by conversion of the enzyme to phosphorylase a (GPa) by phosphorylation of a single serine residue at position 14 in the primary sequence. GPa is almost fully active in the absence of AMP although substrate binding remains cooperative¹⁰. Conversion to the R state is complete in the presence of AMP as both the activator and substrates are bound non-cooperatively¹¹. Phosphorylation increases the affinity of the enzyme for AMP by nearly two orders of magnitude¹². Inhibitors such as glucose, glucose-6-phosphate (G6P) and purines are bound cooperatively by both forms of the enzyme.

The goal of many structural and enzymatic studies of this system over the past forty years has been to understand the molecular basis of phospho-activation and its effects on intersite cooperativity. Here, we show that the predominant structural changes associated with phosphorylation occur at the subunit interface as a result of a conformational transition at the amino terminus. Residues surrounding the phosphorylation site that

participate in intrasubunit interactions in the dephosphorylated state are observed to adopt alternate side-chain conformations in the phosphorylated state, enabling them to form intersubunit interactions. These structural changes account for the strong and non-cooperative binding of AMP by GPa. We interpret the conformational transition induced by phosphorylation of glucose-inhibited GP as an intermediate conformational state in the activation pathway of the enzyme.

GPa and GPb crystallize in isomorphous tetragonal unit cells under similar conditions^{13,14}. Crystals of GPb are obtained in the presence of IMP a weak activator¹⁵ that binds to the nucleotide effector site and confers activity without the increase in substrate affinity observed with AMP. GPa crystals can be grown only in the presence of glucose which promotes the T state by binding and restricting access to the catalytic site¹⁶. Molecular structures of GPa and GPb have been determined independently^{7,9,17-20} and refined atomic models have been obtained with resolutions corresponding to 1.9 Å (GPb) (K.R.A. *et al.*, unpublished results) and 2.1 Å (GPa) (S.R.S. *et al.*, unpublished results; Table 1). Despite the difference in phosphorylation, GPa and GPb adopt similar conformations in the tetragonal crystals and may be described in terms of the same molecular architecture (Fig. 1).

Structure of the subunit interface

The two subunits of the functional dimer are related by the crystallographic twofold axis of symmetry that passes from the catalytic face of the dimer, which contains the substrate-binding sites, to the regulatory face where the phosphoserine residue and the AMP sites are located. The diad axis passes the tower helix $\alpha 7$ (residues 261-277; Fig. 1a) that is linked to a reverse turn (residues 282-285, connecting $\alpha 7$ to $\alpha 8$) which prevents substrates entering the T-state active site²¹. The catalytic sites of the subunits are linked by contacts between residues of the symmetry-related tower helices, which interdigitate to form hydrogen bonds and van der Waals contacts (Table 2). Residues 287-313 form an α -helix ($\alpha 8$) which crosses the molecule, linking the catalytic face to the regulatory face and terminates at the AMP-binding site. On this side of the dimer, the diad axis passes between the $\alpha 2$ helices (residues 50-75) of the subunits (Fig. 1b). The phosphorylation site is located between the $\alpha 2$ helix and the cap (residues 40-45). The $\alpha 2$ helices link the phosphorylation sites of the two subunits. The two subunits enclose a large solvent-filled cavity²² which is girded on each side by β -strands from the amino-terminal domains of both subunits.

|| To whom correspondence should be addressed.

Table 1 Comparison of GP α and GP β

Crystal data: space group P4 ₃ 2 ₁ 2									
	Unit cell (Å)			Resolution (Å)	Reflections	R _{sym}			
	a	c							
GP α	128.4	116.4		2.1	35,231	0.05			
GP β	128.5	116.3		1.9	71,492	0.09			
Refinement parameters									
	Number of protein atoms	Number of water molecules	Δ bond (Å)	Δ angle (Å)	Resolution (Å)	Number of reflections	R _{cryst}		
GP α	6624	770	0.015	0.050	6.0–2.1	34,792	0.15		
GP β	6706	693	0.023	0.054	∞ –1.9	61,344	0.18		
r.m.s. difference in C α positions*									
Residue range	Before least squares (Å)	After translation (Å)	After rotation (Å)	Translation (Å)			Rotation (°)		
				Δx	Δy	Δz	Φ	Ψ	χ
23–834 (D)	0.74	0.60	0.62	–0.46	0.34	0.04	–73.3	133.9	0.53
23–834	0.59	0.48	0.42	–0.31	–0.06	0.31	–35.8	108.5	0.70
23–130	0.73	0.65	0.50	–1.46	–0.83	0.92	–0.2	147.0	2.05
23–314	0.64	0.58	0.48	–0.58	–0.26	0.55	–6.1	139.4	0.86
325–484	0.52	0.36	0.27	–0.43	–0.06	0.50	–35.0	105.3	1.17
485–834	0.57	0.38	0.34	–0.24	0.03	0.11	–45.2	77.7	0.75
Disordered segments in GP α and GP β †									
Residues	GP α	GP β							
1–4	D	D							
5–16	O	D							
17–18	D	D							
19–21	D	O							
250–260	D	O							
314–325	D	O							
839–841	D	O							

Crystallization conditions: GP β : 20–30 mg GP β ml^{–1}, 10 mM *N*, *N*-bis (2-hydroxyethyl)-2-aminoethane sulphuric acid (BES), 10 mM magnesium acetate, 2 mM IMP, 3 mM DTT, 0.1 mM EDTA, pH 6.7. Under these conditions, IMP is bound weakly and the conformation of the enzyme is essentially identical to nucleotide-free GP β ; GP α : conditions for crystallization of GP α were identical to those for GP β except that 50 mM α -D-glucose was substituted for IMP. Crystallographic definitions: $R_{\text{sym}} = \sum_i \sum_h |\bar{I}(h) - I_i(h)| / \sum_i \sum_h I_i(h)$; where $\bar{I}(h)$ and $I_i(h)$ are the mean and the i th measurement of reflection h , respectively. $R_{\text{cryst}} = \sum_h |F_o - F_c| / \sum_h F_o$ where F_o and F_c are the observed and calculated structure factors of reflection h . B factors are the temperature factors obtained from crystallographic refinement.

* Coordinates of α -carbon atoms of the GP β dimer were first superimposed by the least squares method²³ on to the corresponding C α coordinates of GP α . Results of this calculation are listed under 23–834(D). The GP β coordinates transformed in this way served as the basic set for the subsequent superpositions. The translation component was applied first followed by a rotation to give the final operation, $x'_b = R(x_b + T)$ where R is the rotation matrix, T the translation vector and x'_b and x_b are respectively the transformed and untransformed coordinates of GP β .

† Disordered, D; ordered, O. In GP β , regions 250–260, 314–325 and 839–841 represent some of the least well-ordered regions. The residues were located during the later stages of refinement with inclusion of all data, including low-resolution terms, together with solvent correction factors. The electron density of these residues in the final $2F_o - F_c$ maps was satisfactory although the atomic temperature factors in these regions are higher than average and some atoms have B factors greater than 100. The distinction ordered/disordered is qualitative and may reflect the different resolution ranges used in the two refinements. Residues 1–18 remain disordered in GP β even after refinement. Despite the differences in refinement strategies, the overall structures of GP β and GP α are remarkably similar and differences, where observed, are evident in the electron density maps.

Global differences between GP α and GP β

More than 95% of the polypeptide chain is well ordered in GP α and GP β . In the presence of glucose, the $b \rightarrow a$ transition is accompanied by correlated structural changes at the subunit interface; the essential features of the secondary and tertiary structure of phosphorylase are unchanged. Global differences between GP α and GP β have been assessed by least-squares superpositions²³ of α -carbon atoms of GP β with corresponding atoms in GP α dimers, subunits and subunit domains (Table 1). After superposition of the dimers, the $b \rightarrow a$ transformation involves a slight translation (0.57 Å) and rotation ($<1^\circ$) with the result that the subunits are slightly closer to the twofold axis in GP α . The axis of rotation is perpendicular to the molecular diad and passes approximately through the central interface β -strand (184–194). The differences occur predominantly in the subunit interface as there is a considerably larger translation of

1.9 Å and rotation of 2.0° when only the first 130 residues are included in the superposition. The root mean square (r.m.s.) difference in corresponding α -carbon positions is 0.62 Å after the phosphorylase dimers are superposed, which is somewhat greater than the estimated accuracy (0.2 Å) of the refined atomic positions²⁴ in both structures. There is no evidence for any relative movement between the catalytic and regulatory domains. The mean temperature factor for the 791 ordered residues common to GP α and GP β is 25 Å^2 for GP β and 17 Å^2 for GP α . The variation of temperature factors with sequence position is highly correlated for the two structures despite the difference in absolute values (because only the strongest data was used in the refinement of GP α). Relative differences between GP α and GP β subunits in residue positions are colour coded in Fig. 1b. The largest displacements occur at the subunit interface and are a direct consequence of the order/disorder transitions due to serine phosphorylation.

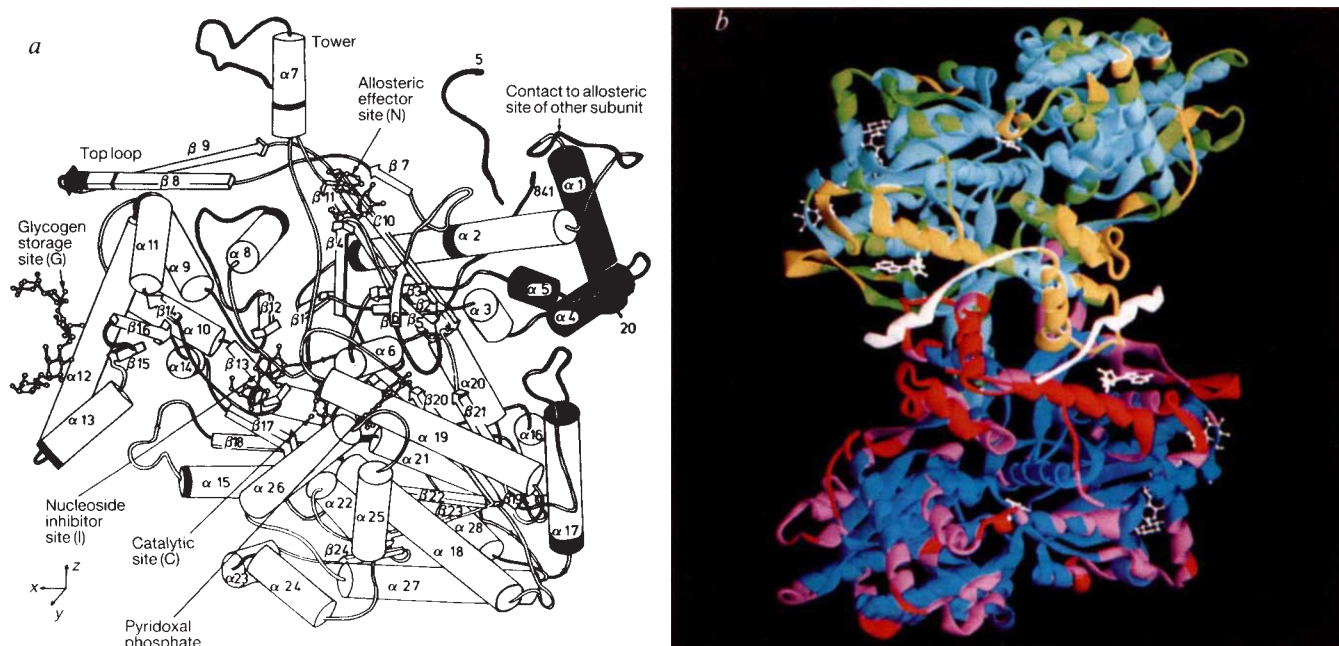


Fig. 1 *a*, Schematic diagram of the phosphorylase subunit viewed down the crystallographic y axis. α -helices and β -strands are shown as cylinders and arrows respectively and are numbered as described⁹. Regions where C_{α} positions differ by more than 0.6 Å between GPb and GPa are shown black. In GPa the N-terminal tail (length 5–16 shown here) displaces the C-terminal region of GPb (residues 837–841). Each subunit is composed of two major domains. Domain 1, the regulatory domain, consists of the first 484 residues (from the N terminus up to and including strand β 18) and is organised about a mostly parallel, nine-stranded, β -sheet core. This domain is subdivided (with the division after helix α 8) into interface (residues 1–315) and glycogen-binding (residues 316–484) subdomains. Domain 2 (residues 485 to the C terminus) contains a classical nucleotide-binding fold which is flanked by two layers of helices. The catalytic site is located between domains 1 and 2. The essential cofactor pyridoxal phosphate is linked by a Schiff base to Lys 680 in helix α 21. *b*, Ribbon schematic diagram of the phosphorylase *a* dimer regulatory face viewed along the twofold rotation axis (the $x = y$ axis of the tetragonal unit cell and -135° to the view shown in *a*). Each subunit is colour-coded to show the difference in C_{α} positions between GPb and GPa. Shifts 0–0.4 Å: blue (bottom subunit) and cyan (top subunit); 0.4–0.6 Å: pink (bottom subunit) and green (top subunit); >0.6 Å: red (bottom subunit) and yellow (top subunit). The N-terminal tail containing the phosphorylation site at Ser 14 is shown in white in both subunits. Structural differences in residues 208–211 at the end of a β -ribbon (top loop; β 8, β 9) and at discrete sites within helical segments at the domain interface involve crystal contacts between phosphorylase dimers and thus reflect changes in crystal packing rather than direct effects of phosphorylation. AMP bound at the allosteric effector site, maltoheptaose bound at the glycogen storage site and G1P at the catalytic site are shown at top right, right and centre, respectively, of the bottom subunit. Ribbon drawing produced with software written by W. M. Carson.

Order/disorder transitions

The polypeptide segment which contains the phosphorylation site (residues 5–16) is disordered in GPb, but folds into a distorted 3_{10} helix in GPa (residues 14–16 are extended). The helix is bound within a broad cavity at the subunit interface (Figs 1*b*, 2 and 3) through an extensive network of hydrogen bonds and van der Waals interactions (Table 2; Fig. 2). Contacts between polar groups result in the formation of eight inter-subunit and eight intrasubunit hydrogen bonds. Phosphoserine 14 (Ser-P14) is bound directly to the subunit interface by two ion-pair interactions, one with Arg 69 of the α 2 helix and the second with Arg 43' in the cap of the twofold-related subunit (residues of this subunit are designated by primes)^{19,22}. These residues remain accessible to solvent. The phosphoserine group also interacts with the main-chain amide of Val 15 and, through a water molecule, with the side chain of Gln 72 (Table 2, Fig. 3).

Hydrophobic interactions are important in stabilizing the N-terminal 3_{10} helix at the subunit interface. Isoleucine 13 and Val 15 are completely buried in pockets lined with hydrophobic residues (Table 2; Fig. 2). Part of the hydrophobic binding site for these residues is formed by the methylene side-chain atoms of Arg 43' and Arg 69, the phosphate-binding residues. Tyrosine 51' and Phe 53' from the α 2' helix contribute to the bottom of the hydrophobic binding surface of Ile 13 and Val 15 is bound between the α 1' and α 2 helices. The contact surface between the reordered N terminus and its binding site, calculated from the difference between the solvent-accessible surface area²⁵ of the isolated N terminus (residues 5–16) and that of the corresponding segment in GPa, extends over 1,400 Å² and constitutes

more than 25% of the interaction surface between the diad-related GPa subunits. The molecular surface involved in inter-subunit contacts is not extensive in GPb, comprising only 7% (2,294 Å²) of the total solvent-accessible surface area of the isolated monomer, but increases to 10% (3,450 Å²) on phosphorylation. Non-polar groups account for 60% of this buried surface.

On binding to the subunit interface, the phosphorylated N terminus displaces four C-terminal residues, 838–841, which become disordered in GPa (residue 842 is disordered in both GPb and GPa; Fig. 3*a*). The position occupied by Glu 839 in GPb is taken up by Arg 16 in GPa. The displacement of the C terminus disrupts an ion-pair bond between Asp 838 and His 36'. The histidine residue rotates by 120° about the C_{α} -C β bond and forms new hydrogen bonds with the carbonyl oxygen of Lys 41' and the backbone amide of the phosphoserine residue in the opposite subunit (Figs 3, 4; Table 2). Rotation of His 36 also removes a steric barrier to the binding of Val 15 to its hydrophobic pocket between Val 64 and Ile 68 (Fig. 4*b*). Thus, by a rotation about the His 36 C_{α} -C β bond, an interaction between the C- and N-terminal domains is broken and new contacts between subunits are made.

The conformation of the first ordered residues of the α 1 helix indicate that the N terminus of GPb is displaced from the subunit interface. In GPa, Glu 22 is oriented toward Arg 16, whereas the five residues that complete the covalent link between the α 1 helix and the N terminus are disordered. In GPb, the α 1 helix is well ordered from residue 19 and thus contains an extra turn of ordered polypeptide chain. However, residue 19

Fig. 2 A stereo diagram of Ser14 phosphate-binding site in GP_a. The view is approximately along the crystallographic *a* axis, 90° from that in Fig. 1*a*, facing towards the side of the subunit. Main-chain atoms are shown for residues 5–15, 68–72, 36'–43' and 115'–116' and the side chains of Arg 10, Ile 13, Ser-phosphate 14, Val 15, Ile 68, Arg 69, Gln 72, His 36', Phe 37', Asp 42', Arg 43', Tyr 51' and Ile 115'. The position of an ordered water molecule (W) present in GP_a is marked with a +. Carbon atoms are coloured magenta in the 5–15 segment and green elsewhere. Oxygen and nitrogen atoms are coloured red and cyan, respectively, for all residues. The diagram illustrates the hydrogen-bonding network surrounding the phosphoserine group involving Arg 69, Gln 72, an ordered water molecule, Asp 42' and Arg 43'. The hydrogen-bond contact between Arg 10 and the main-chain carbonyl of 116' is shown.

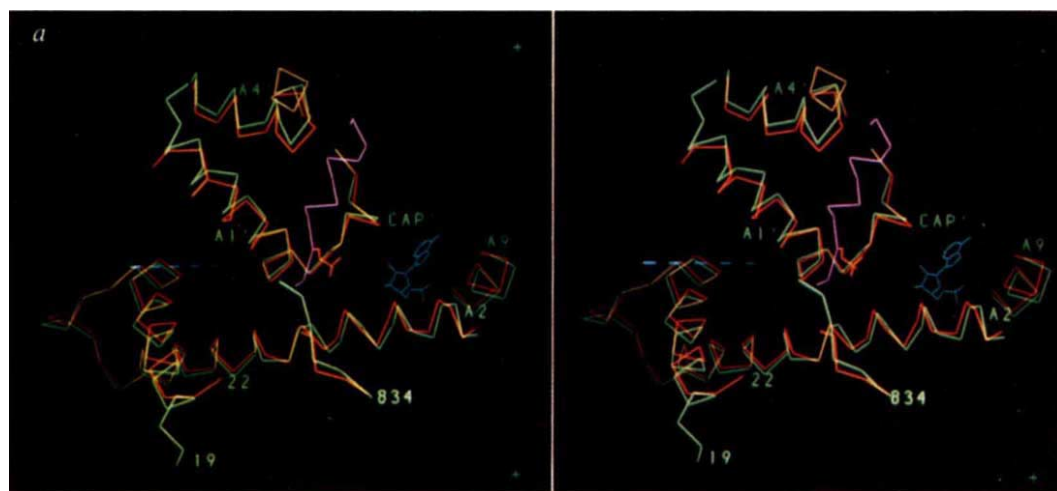
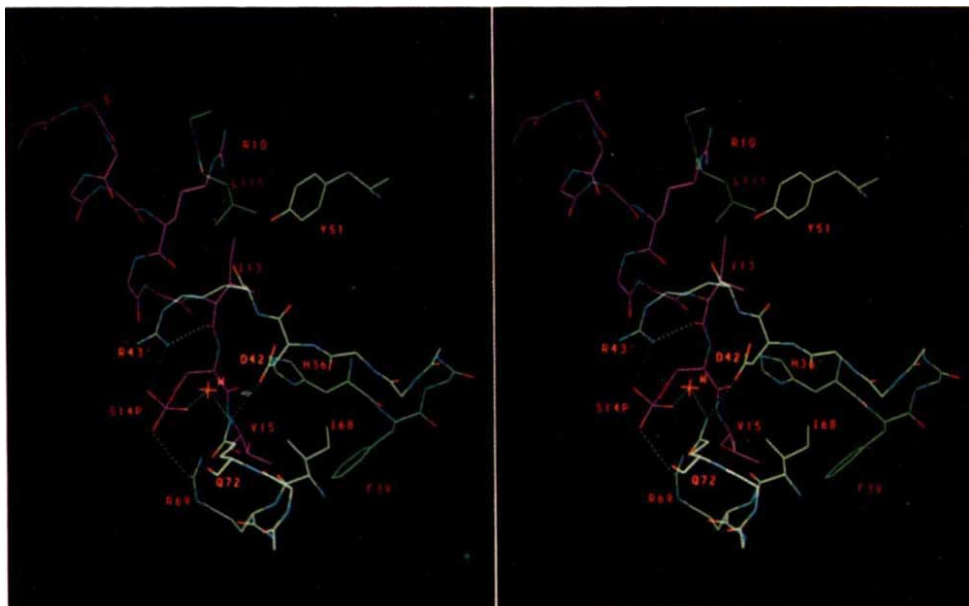


Fig. 3 *a*, Conformational changes at the subunit interface and phosphorylation site: Ca tracing for GP_b is shown in green; for GP_a in red. AMP as bound in GP_b is cyan. The view is rotated slightly from that in Fig. 2. The crystallographic diad axis is shown as a blue dashed line. Movements of the $\alpha 1$, $\alpha 2$, $\alpha 1'$ and $\alpha 4'$ helices towards the opposite subunit are apparent in the *b* $\rightarrow$ *a* transition; the ordered N terminal tail (residues 5–16, shown here in magenta) in GP_a displaces the C terminus of GP_b. *b*, Detail of the phosphoserine site in GP_b shows His 36'/Asp 838 and Arg 43'/Leu 115' hydrogen bonds. *c*, These are broken in GP_a and replaced with Arg 43'/phosphoserine/Arg 69/Gln 72/Asp 42' intersubunit hydrogen-bond network and Arg 10/Leu 115' intersubunit hydrogen bond.

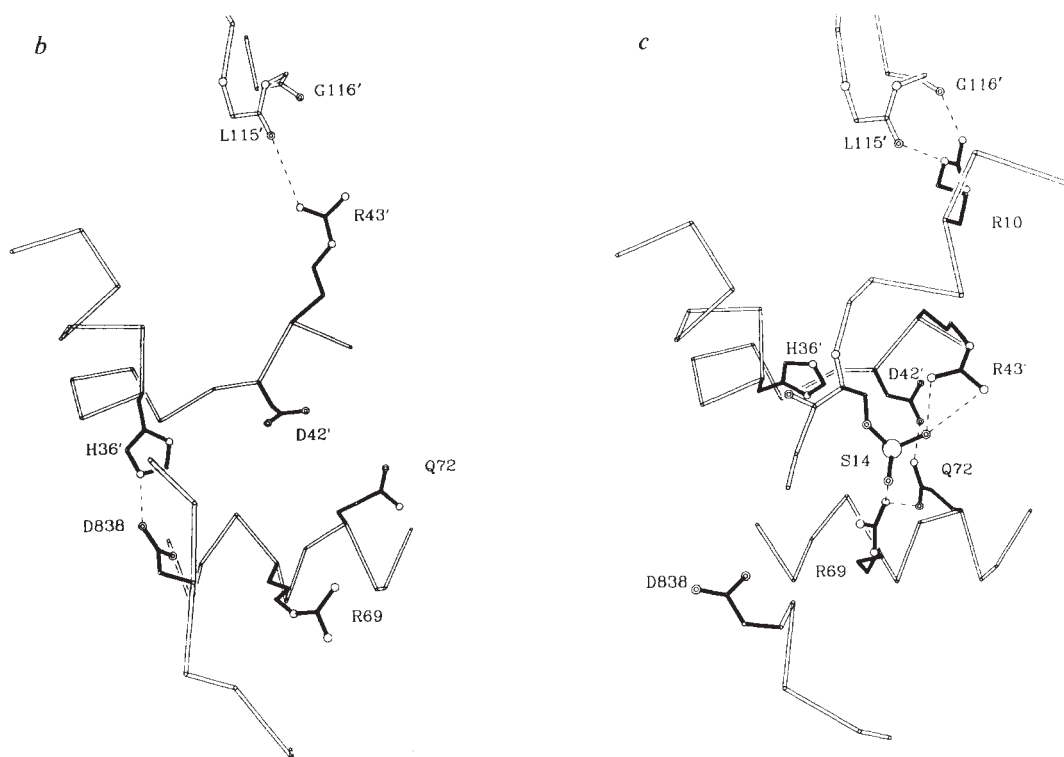


Table 2 Inter and intramolecular contacts of the N-terminal tail in GP α and comparison of intersubunit hydrogen bonds* between GP α and GP β GP α : N-terminal tail only

Residue	Intersubunit van der Waals contacts	Intersubunit hydrogen bonds			Intrasubunit hydrogen bonds		
K9	Y113, Q114'	(N ζ	O	Y113')	O	N	Q12
R10	R43', Y51', L115'	N η 2	O	G116'	O	N	I13
	G116, L117'	N ϵ	O	G116'			
K11	R43'						
Q12	K28', N32', Q114'	O	N δ 2	N32'	O	N η 2	R16
	L115'	N ϵ 2	O	Q114'			
I13	N32', L35', H36'	O	N η 1	R43'			
	R43', L115'						
S-P14	N32', H36', R43'	N	N ϵ 2	H36'	O ϵ 2	N η 2	R69
		O	N δ 2	N32'	O ϵ 3	N η 1	R69
		O ϵ 1	N η 1	R43'	O ϵ 3...	WAT...N ϵ 2	Q72
					O ϵ 3	N	V15
V15	H36'						
R16					N η 2	O	Q12

Total intersubunit hydrogen bonds: 8

Intersubunit hydrogen bonds (excluding N-terminal tail)

GP β						GP α					
(O	F37	N η 1	R60'			O	F37	N η 1	R60'		
O ϵ 1	Q72...	WAT...		O δ 1	N44'	N ϵ 2	Q72	O δ 1	D42'		
O ϵ 1	E178	N δ 2	N250'								
						O δ 1	D181	N ζ	K247'		
O δ 1	D181...	WAT...		O	A248'	O δ 1	D181...	WAT...		O	A248'
				N η 1	R269'					N η 1	R269'
N η 1	R193	O	V40'			N η 1	R193	O	V40'		
N η 2	R193	O	L39'			N η 2	R193	O	L39'		
O ϵ 1	E195	N ζ	K41'			O ϵ 1	E195	N ζ	K41'		
O η	Y262...	WAT...		N	D180'	O η	Y262...	WAT...	N		D180'
O δ 1	N270	N ϵ	R277'			(O δ 1	N270	N η 1	R277'		
N δ 2	N270	O δ 1	N274'			N δ 2	N270	O δ 1	N274'		
O δ 1	D838	N ϵ 2	H36'								
Total: 11						Total: 10					

* Cut-off distance for hydrogen bonds is 3.4 Å. Interactions in brackets have a separation of atoms between 3.4–3.6 Å. Residues are identified by the one-letter code. Ordered water-molecules are designated by WAT.

is turned toward the α 4 helix and away from the phosphoserine binding site (Fig. 3a) suggesting that the N terminus is swung away from the interface.

Intersubunit contacts tighten

Phosphorylation of Ser 14 triggers the formation of intersubunit contacts involving residues remote from the phosphorylation site. In GP β , Arg 43' forms a weak hydrogen bond with the main-chain carbonyl oxygen of residue 115' at the C terminus of the α 4' helix in its own subunit (Fig. 3b). This contact is broken in GP α when both Arg 43' and Arg 69 rotate to form ion-pair bonds with Ser-P14 (Fig. 3c). The new ionic interaction pulls the cap toward the α 2 helix of the opposite subunit. Following the motion of the cap, the cluster formed by α 4'/ α 5' helix corner and the α 1' helix rotates toward the opposite subunit (Fig. 3a) to form contacts with the N-terminal tail. Intrasubunit contacts between 115' and 43' in GP β are replaced by intersubunit hydrogen bonds between Arg 10 and the carbonyl oxygen atoms of residues 115' and 116'. In both the *a* and *b* states, an arginine residue interacts with the C terminus of an α -helix.

Rotation of Arg 43' also creates part of the hydrophobic binding pocket for Ile 13. As the α 1' helix shifts along its axis toward α 2, a hydrogen bond is formed between Asp 42' and Gln 72 and van der Waals contacts between Phe 37', Val 40' and Val 64 are broken. The new interactions, together with contacts between Gln 72 and Arg 69 and Arg 69 with Ser-P14, complete a hydrogen-bond network linking the subunits together (Fig. 3c). The increase in intersubunit contact surface (see above) indicates that the tighter association is global.

Alternative binding modes at the AMP site

The allosteric effectors AMP and G6P, like the phosphoserine residue, are bound between the α 2 helix of one subunit and the cap of the opposite subunit. However, slightly different side-chain interactions are involved in effector binding. The *b* $\rightarrow$ *a* transition decreases the separation between the α 2 helix and the cap and induces the α 2 helix to contract. Both of these structural changes create a stronger binding site for the activator AMP and a weaker binding site for the inhibitor G6P. In GP β the α 2 helix is interrupted by a turn of 3_{10} helix between the amide group of residues 64 and the carbonyl oxygen of residue 61 (Fig. 4). The turn is partly stabilized by a hydrogen bond between His 62 and the carbonyl oxygen of Thr 58. Histidine 62 is fixed by an ionic interaction with Glu 26. The side chain of His 36' packs with Val 64 and Ile 68 and may act as a wedge between adjacent turns of the α 2 helix. After the 3_{10} turn, the helix is unwound at residues 65 and 66 and this extends the α 2 helix of GP β about 0.8 Å towards the AMP site relative to GP α .

The formation of an ion pair between Arg 69 and Ser-P14 in GP α exerts a restoring force on the α 2 helix, promoting the formation of α -helical hydrogen bonds at the 3_{10} kink. Model building shows that the contraction of the α 2 helix allows shorter ionic contacts between Arg 69 and the phosphoserine group than would be possible in the presence of the kink. The interactions stabilizing the 3_{10} turn are absent in GP α as the shift of the α 1 helix disrupts the Glu 26–His 62 hydrogen bond, leaving His 62 unconstrained and destabilizing its interaction with Thr 58. Displacement of His 36' (see above) also promotes the restoration of helical contacts between residues 64 and 68. The

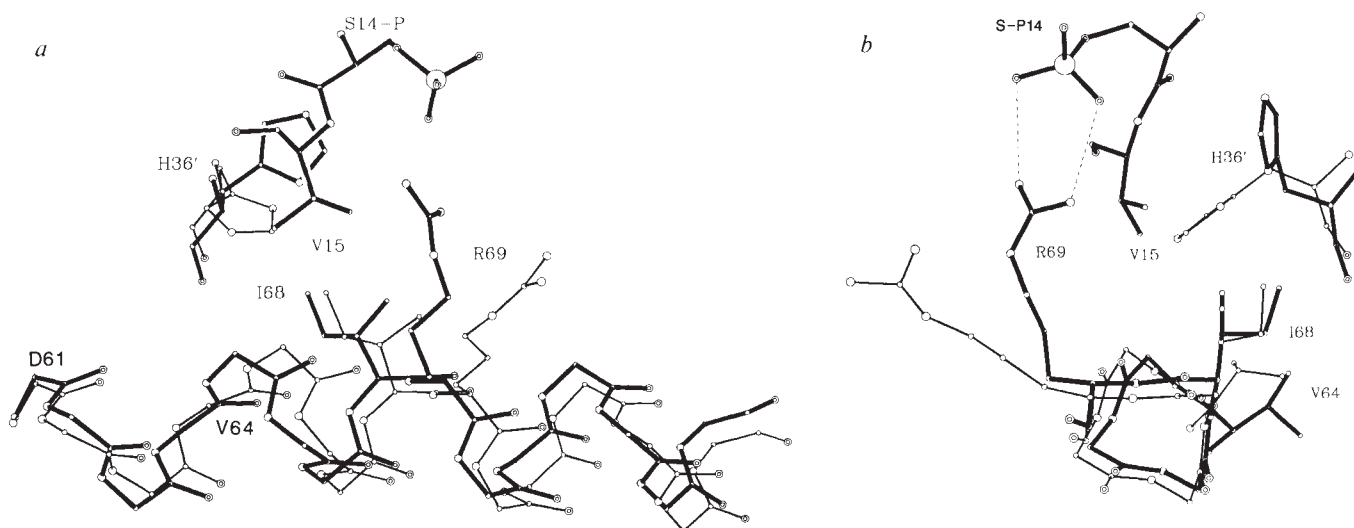


Fig. 4 Helix $\alpha 2$ is unwound at residues 61–64 in *GPb* (thin lines) and annealed in *GPa* (thick lines). *a*, View parallel to the helix axis. *b*, View down helix with main-chain atoms of residues 63–69 shown. In *GPb* His 36' packs against Val 64 and Ile 68. In *GPa* His 36' swings out as Arg 69 makes contact with Ser P14 and Val 15 packs against the non-polar side chains. These are some of the interactions that seem to favour unwinding of helix $\alpha 2$ (for other interactions see text).

consequent realignment of the $\alpha 2$ helix with the cap of the nucleotide-binding site is responsible for alternate effector binding modes.

Crystallographic analysis of complexes of AMP with *GPb* (refs 9, 26, 27) and *GPa* (refs 28, 29) reveal differences both in the conformation of the ligand and in its interaction with binding-site residues (Fig. 5*a, b*). In *GPb*, the side chain of Tyr 75 is stacked over the purine ring of AMP at a distance of 4 Å. The contraction of the $\alpha 2$ helix in *GPa* is accompanied by a 120° rotation of the tyrosine ring about the $C\alpha$ - $C\beta$ bond. The structural change at the binding site is reflected in an altered binding mode of the ligand. Relative to its orientation in *GPb*, AMP bound to *GPa* is rotated by 45° about an axis perpendicular to the ribose plane and translated by 1.0 Å toward the interior of the protein. This allows a close, 3.4 Å stacking interaction to form between Tyr 75 and the adenine ring and promotes additional contacts between the ribosyl group and phosphorylase side chains. Asn 44', Asp 42' and Gln 71 all form hydrogen bonds with the ligand in *GPa* (Fig. 5*b*), whereas in *GPb* only the contact with Asp 42' is made. In addition, AMP buries more solvent-accessible surface area (by 20 Å²) than in the complex with *GPb*. The phosphate group of AMP forms ion-pair hydrogen bonds with Arg 309 (after rotation of the side chain) and Arg 310 in complexes with both *GPa* and *GPb*, but the phosphate itself is in a different position in the two complexes.

Although AMP interacts more strongly with *GPa*, fewer protein side chains make contact with the allosteric inhibitor G6P (R.J.F. *et al.*, unpublished results). Glucose-6-phosphate binds to *GPb* with the hexose ring between the $\alpha 2$ helix, the $\beta 7$ strand and the cap of the opposite subunit²⁶ (Fig. 5*a*). The phosphate-binding site is 2 Å from that of the AMP phosphate site and the phosphate forms ion pairs with Arg 309 and 310 as does AMP, but it also interacts with Arg 242 from the underlying β -sheet. In contrast, although the phosphate group of G6P is bound tightly by the arginine cluster in *GPa*, the hexose ring is completely disordered. In response to the conformational changes induced by phosphorylation, the side chain of Gln 71 rotates into the hexose-binding site. The rotation of Gln 71 may be a consequence of the contraction of the $\alpha 2$ helix and the altered positions of Gln 72 and Asp 42'. The movements of these residues serve to inhibit G6P binding but promote the binding of AMP.

Conformational differences between *GPb* and *GPa* at the catalytic site are mostly due to the presence of glucose in *GPa*,

which also prevents the propagation of conformational changes from the subunit interface to the catalytic site²¹. The cofactor, pyridoxal phosphate, is almost identical in the two structures although there are some small differences in some of the contacts³⁰. The changes at the catalytic site will be described in detail elsewhere.

Mechanism of phosphorylation

Activation of *GPb* by phosphorylation is achieved with an allosteric mechanism in which the functional conformation of the enzyme is stabilized by specific interactions with the phosphorylated N-terminal helix. The involvement of the entire N terminus in subunit contacts explains why inorganic phosphate does not bind to the phosphoserine locus in *GPb*²⁶. Because the intersubunit binding site of the N and C terminals partially overlap, attachment of the N terminus requires release of the C terminus from the surface of the enzyme. The N-terminal helix is thus an effector of the R state whereas the C terminus promotes the T state partly by acting as a competitive effector.

Studies of phosphorylase *b'*, a proteolytic product obtained by tryptic cleavage at Arg 16, show that the N terminus is essential for subunit cooperativity and, in *GPb*, stabilizes the T state. Phosphorylase *b'* can be activated either by AMP or a phosphorylated tetradecapeptide corresponding to residues 3–16, but exhibits no heterotropic cooperativity^{31,32}. Removal of the N terminus allows the subunit to exhibit R-state-like properties. For example, the affinity of *GPb'* for AMP is approximately sixfold greater than that of native *GPb* at low phosphate concentrations. Although the nucleotide effector promotes an active catalytic site conformation, the covalently linked N terminus is required to convey conformational signals to the diad-related subunit. Cross-linking and spectroscopic studies of the native enzyme indicate that AMP, unlike phosphorylation, does not induce the N terminus to assume an ordered state³³.

The close intersubunit contacts promoted by phosphorylation also increase the affinity of the enzyme for AMP. The dissociation constant for AMP decreases from 100–200 μ M for *GPb* (ref. 34) to 2 μ M in *GPa* (ref. 11). The change in affinity results from a transition between two different binding modes which can be distinguished kinetically³⁵ and by fluorescence³⁶. The structural results reported here indicate that these binding modes correspond to alternative orientations of the bound effector. The mode used in *GPa* allows a more extensive interaction between

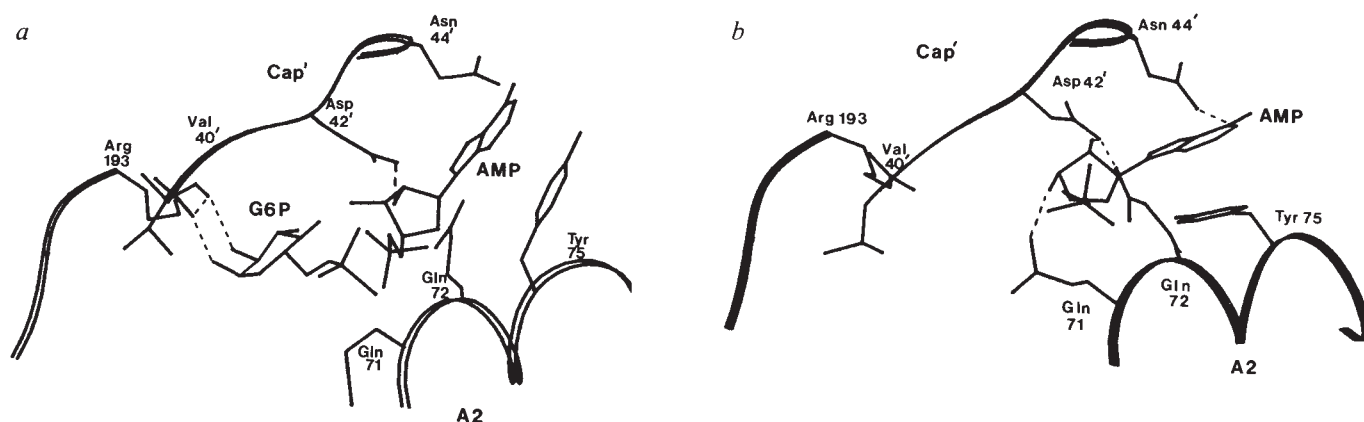


Fig. 5 Identical views showing the alternate binding modes of AMP. *a*, GPb: Crystals of GPb were soaked in a solution containing 100 mM AMP or 100 mM G6P in the crystallization buffer as described^{26,27}. Coordinates and shifts in protein atoms were determined from difference Fourier syntheses. The adenine of AMP is stacked against Tyr 75 and O2' hydroxyl of the ribose is hydrogen bonded to Asp 42'. The position of G6P is also shown although AMP and G6P do not bind simultaneously. *b*, GPa: Crystals of GPa were soaked in a solution containing 5 mM AMP. Positional coordinates of the complexed enzyme were refined by crystallographic least-squares methods as described²⁹. The contacts are more extensive and include more favourable stacking with Tyr 75 and the O2' to Asp 42' and O3' to Gln 71. Gln 71 blocks G6P binding in GPa. Contacts to the phosphate moieties are not shown. AMP bound to either GPb or GPa has a similar anti conformation about the N-glycosidic link. The ribose adopts a C3' endo conformation in GPb and a C2' endo conformation in GPa. Thus there is a large difference in the torsion angles about the C4'-C5' bonds: in GPb the conformation is *gauche*; in GPa it is *trans*.

the AMP and the binding-site residues. The orientation of AMP observed in GPb corresponds to a low-affinity complex which is similar to that observed for the interaction of ATP, a T-state effector, to GPa (ref. 29). Likewise G6P is complementary to GPb but not GPa. Glucose-6-phosphate is a powerful inhibitor of GPb (ref. 37) but only inhibits GPa under extreme conditions (ref. 38).

The work presented here characterizes the essential T→R transition at Ser-P14 and allosteric effector site in GP. The structure of GPa complexed with glucose represents a conformational intermediate in the T→R transition in which the regulatory apparatus at the N terminus has been decoupled from the catalytic site by glucose. Structural studies on complexes of both GPb (refs 39, 40) and GPa (refs 28, 41, 42, E.J.G. *et al.*, unpublished results) with substrates, R-state inhibitors and activators have revealed additional conformational changes that lead to increased access to the catalytic site and rearrangements throughout the subunit and domain interfaces. Work is now in progress to characterize more highly activated ligand complexes and alternative crystal forms favouring the R state in both GPb and GPa.

The similarity between the phosphorylation site in GP and those of other enzymes subject to phosphoregulation by ligand dependent kinases^{43,44} may reflect a common structural mechanism of regulation.

The phosphorylation sites in these proteins, like that in GP, are flanked on their N-terminal side by a cluster of basic residues. These residues are required for kinase recognition, but may also play a role in conformational regulation. In phosphorylase, activation requires the N terminus to fold into a compact helical structure that binds to the subunit interface. In GPb, this would result in the unfavourable opposition of positive charges (Lys 9, Arg 10, Lys 11, Arg 16, Arg 69 and Arg 43') both within the helix and in the binding site. Phosphorylation of Ser 14 induces the N terminus to fold into the required helical conformation by compensating for the positive charge built up by the cluster of basic residues. The basic cluster is thus an essential feature of the regulatory apparatus in the sense that phosphorylation triggers a functionally active conformation of the N-terminal 'effector'. Similar sequences may play similar roles in other enzymes subject to phospho-regulation by cyclic AMP- or Ca²⁺-dependent kinases.

This work has been supported by grants from NIH numbers DK31507-06 (S.R.S.); AM26081 (R.J.F.) and the MRC and SERC (L.N.J.). S.R.S. is an assistant investigator of the Howard Hughes Medical Institute. We wish to acknowledge the contributions by our colleagues, past and present, to the structure determinations of GPb and GPa and to Arne Strand and Peter Roome for assistance with figures.

Received 27 June; accepted 12 September 1988.

- Krebs, E. G. in *The Enzymes* 3rd edn Vol. 17 (eds Boyer, P. D. & Krebs, E. G.) 3-20 (Academic, New York, 1986).
- Cohen, P. *Eur. J. Biochem.* **151**, 439-448 (1985).
- Krebs, E. B. & Fischer, E. H. *Biochem. biophys. Acta* **20**, 150-157 (1956).
- Graves, D. J. & Wang, J. H. in *The Enzymes* 3rd edn Vol. 7 (ed. Boyer, P. D.) 435-482 (Academic, New York, 1972).
- Fletterick, R. J. & Madsen, N. B. *A. Rev. Biochem.* **49**, 31-61 (1980).
- Dombardi, V. *Int. J. Biochem.* **13**, 125-139 (1981).
- Fletterick, R. J. & Sprang, S. R. *Acct Chem. Res.* **15**, 361-369 (1982).
- Madsen, N. B. in *The Enzymes* 3rd edn Vol. 17 (eds Boyer, P. D. & Krebs, E. G.) 366-394 (Academic, New York, 1986).
- Johnson, L. N. *et al.* in *Allosteric Enzymes* (ed. Herve, G., CRC in the press).
- Metzger, B. E., Glaser, L. & Helmreich, E. *Biochemistry* **7**, 2021-2036 (1968).
- Engers, H. D., Schechosky, S. & Madsen, N. B. *Can. J. Biochem.* **48**, 746-754 (1970).
- Helmreich, E., Michaelides, M. C. & Cori, C. F. *Biochemistry* **6**, 3695-3710 (1967).
- Johnson, L. N., Madsen, N. B., Mosley, J. & Wilson, K. S. *J. molec. Biol.* **90**, 703-717 (1974).
- Fletterick, R. J., Sygusch, J., Semple, H., Madsen, N. B. & Johnson, L. N. *J. molec. Biol.* **103**, 1-13 (1976).
- Black, W. J. & Wang, J. H. *J. biol. Chem.* **243**, 5892-5898 (1968).
- Kasvinsky, P. J., Schechosky, S. & Fletterick, R. J. *J. biol. Chem.* **253**, 9102-9106 (1978).
- Fletterick, R. J., Sygusch, J., Semple, H. & Madsen, N. B. *J. biol. Chem.* **251**, 6142-6146 (1976).
- Weber, I. T. *et al. Nature* **274**, 433-437 (1978).
- Sprang, S. R. & Fletterick, R. J. *J. molec. Biol.* **131**, 523-551 (1979).
- Sansom, M. S. P. *et al. J. Mol. Struct.* **123**, 3-25 (1985).
- Sprang, S. R., Goldsmith, E. J., Fletterick, R. J., Withers, S. G. & Madsen, N. B. *Biochemistry* **21**, 5364-5371 (1982).

- Sprang, S. R. & Fletterick, R. J. *Biophys. J.* **32**, 175-192 (1980).
- Hendrickson, W. A. *Acta crystallogr.* **A35**, 158-163 (1979).
- Luzzati, V. *Acta crystallogr.* **5**, 802-810 (1952).
- Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 397-400 (1971).
- Lorek, A. *et al. Biochem. J.* **218**, 45-60 (1984).
- Stura, E. A. *et al. J. molec. Biol.* **170**, 529-565 (1983).
- Kasvinsky, P. J., Madsen, N. B., Sygusch, J. & Fletterick, R. J. *J. biol. Chem.* **254**, 3343-3351 (1978).
- Sprang, S. R., Goldsmith, E. & Fletterick, R. J. *Science* **237**, 1012-1019 (1987).
- Oikonomakos, N. G. *et al. Biochemistry* **26**, 8381-8389 (1987).
- Graves, D. J., Mann, S. A. S., Philip, G. & Oliveira, R. J. *J. biol. Chem.* **243**, 6090-6098 (1968).
- Janski, A. M. & Graves, D. J. *J. biol. Chem.* **254**, 4033-4039 (1979).
- Gusev, N. B., Hajdu, J. & Friedrich, P. *Biochem. biophys. Res. Commun.* **90**, 70-77 (1979).
- Mateo, P. L., Baron, C., Lopez-Mayorga, O., Jimenez, J. S. & Cortijo, M. *J. biol. Chem.* **259**, 9384-9389 (1984).
- Kastenschmidt, L. L., Kastenschmidt, J. & Helmreich, W. *Biochemistry* **1**, 4543-4556 (1968).
- Vanderbunder, B., Dreyfus, M. & Buc, H. *Biochemistry* **17**, 4153-4160 (1978).
- Morgan, H. E. & Parmeggiani, A. J. *J. biol. Chem.* **239**, 2440-2445 (1964).
- Melpidou, A. E. & Oikonomakos, N. G. *FEBS Lett.* **154**, 105-110 (1983).
- Oikonomakos, N. G. *et al. Eur. J. Biochem.* **173**, 569-579 (1988).
- Hajdu, J. *et al. EMBO J.* **6**, 539-545 (1987).
- Madsen, N. B., Kasvinsky, P. J. & Fletterick, R. J. *J. biol. Chem.* **253**, 9097-9101 (1978).
- Withers, S. G., Madsen, N. B., Sprang, S. R. & Fletterick, R. J. *Biochemistry* **21**, 4372-4382 (1982).
- Krebs, E. G. & Beavo, J. A. *A. Rev. Biochem.* **48**, 923-959 (1979).
- Cohen, P. *Proc. R. Soc. Lond. B234*, 115-144 (1988).

Unresponsiveness to a foreign antigen can be caused by self-tolerance

Damir Vidović*† & Polly Matzinger*

* Basel Institute for Immunology, CH-4005 Basel, Switzerland

† Centre d'Immunologie CNRS-INSERM F-13288 Marseille-Luminy, France

In mice, two sets of genes govern the immune response to the synthetic antigen GT. One maps to the major histocompatibility complex and behaves like a typical immune response gene. The second is a background gene encoding a cell surface structure found on B cells. Mice which express, and are therefore tolerant of, one form of this structure do not respond to GT. Thus, tolerance of self generates holes in the T-cell repertoire, partially crippling the immune system.

ALTHOUGH the immune system is very efficient at recognizing foreign antigens, it is not perfect. The immune system has specific 'blind spots' which lead to precisely delineated failures in immune responsiveness^{1,2}. There are three quite different sorts of genes that may influence the presence or absence of such blind spots.

(1) Genes for the antigen receptors themselves. This sort of defect has been demonstrated for both B and T cells. For example strains of mice which lack the antibody variable (V) region gene 'V_{dex}' are unable to respond against the sugar polymer α -1-6 dextran^{3,4} and mice which lack the V α 2 T-cell receptor (TCR) gene⁵, cannot produce cytotoxic T cells to the antigen H-Y.

(2) Genes in the major histocompatibility complex (MHC), or immune response (Ir) genes. Unlike B cells, T cells do not recognize foreign antigen in their native form, but only when displayed on the surface of a cell in association with one of the MHC proteins⁶. When no MHC-antigen complex forms, no T cell can respond. Because the products of some MHC alleles cannot form complexes with some particular antigens⁷, animals bearing these MHC alleles are specifically unresponsive.

(3) Genes encoding normal body components which the immune system ought not to attack. Each individual maintains a state of self-tolerance, preventing destructive immune responses against its own bodily structures⁸. Any foreign molecule sufficiently similar to one of these 'self' structures, would be safe from the immune system.

Benacerraf and McDevitt first suggested self-tolerance of the MHC as the basis for MHC-linked Ir genes¹, but dropped the idea for two reasons. Firstly, responder $\times$ non-responder, (R $\times$ NR)F₁ hybrids should be unresponsive, but theirs responded. Second, they could not immunize responder strains using tissues from non-responders. When the concept was extended to include non-MHC molecules and their associations with MHC products, however^{9,10}, self-tolerance became an inevitable basis for unresponsiveness. Although there have been hints that this third type

of defect may exist¹¹⁻¹³ the evidence has been circumstantial². We present here an extensive analysis of one case in which tolerance of 'self' clearly cripples the response to a foreign antigen.

Ir genes control responses to GT

The immune response of T cells against the synthetic polymer poly(Glu⁵⁰Tyr⁵⁰) (GT) is controlled by the class II genes of the mouse MHC¹⁴. Mice that express the *q*, *s* or *w13* alleles of the *A* locus are responders, whereas (with rare exceptions) mice carrying the *f*, *k*, *j*, *v*, *u*, *b* and *d* alleles are non-responders and, in back-crosses, the response segregates with the R MHC alleles¹⁴. This pattern of responsiveness is compatible with either of the first two classes of defect discussed above. Several exceptional responses have been found, however, which revealed the existence of a second controlling gene not linked to the MHC¹⁴. For example, DBA/2 mice respond to GT, but BALB/c and B10.D2 do not, even though all three strains are identical for the entire set of class II genes. Furthermore the (R $\times$ NR)F₁ hybrids were unresponsive, suggesting a type 3 defect.

Our first step in analysing this response was to determine the number of background genes involved, and whether they were actually unlinked to the MHC.

Mapping the second gene

We crossed the two NR strains B10.D2 and BALB/c with the R strain DBA/2. As all three strains are of the *d* MHC type, only non-MHC immune response genes should segregate in these crosses. The F₁ progeny were all non-responders and the 96 back-cross progeny segregated approximately equally into R and NR categories (Fig. 1), indicating that only one controlling locus segregates in mice bearing B10 and BALB backgrounds.

Using recombinant inbred mice¹⁵ we tentatively mapped the gene in the B6 background onto chromosome 9, distal to *bg1* (ref. 14). This ruled out the known T-cell receptor genes^{16,17}

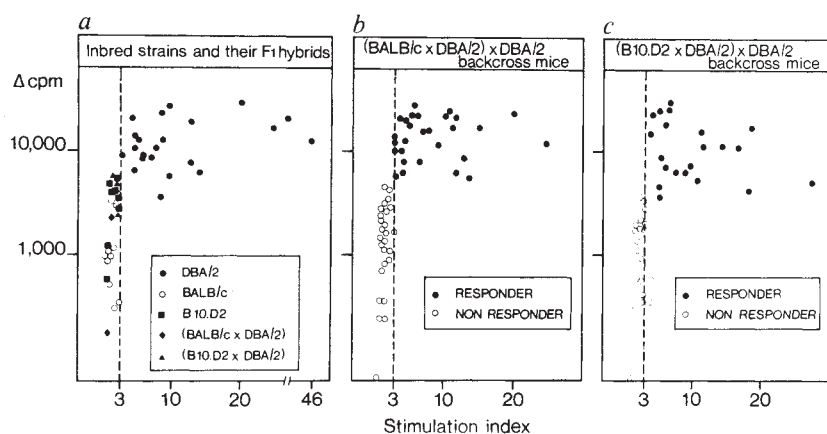


Fig. 1 Responsiveness to GT segregates with a single non-MHC gene. Male or female mice were immunized as previously described¹⁴ and tested when 8–14 weeks old. Briefly, they were injected with 100 μ g GT in complete Freund's at the base of the tail. From 7–8 days later, cells from the draining LN were cultured in microwells at 4×10^5 cells per 0.2-ml well containing a final concentration of 125 μ g ml⁻¹ GT. After three days, cell growth was measured by incorporation of ³H-thymidine (c.p.m. = counts per minute of ³H-thymidine incorporated). Responses are the arithmetic means of triplicate cultures. We have plotted the responses as scatter diagrams of Δ c.p.m. versus stimulation index (SI). Both parameters are valuable in distinguishing between responder and non-responder mice.

Table 1 BALB/c LPS blasts and GT can be interchangeably used for priming responder DBA/2 mice

<i>In vivo</i> priming	<i>In vitro</i> boost	Responses (c.p.m.)	
		Expt 1	Expt 2
LPS blasts			
—	—	1,245	2,796
—	GT	2,165	2,492
DBA/2	—	1,435	4,072
DBA/2	GT	1,934	5,244
BALB/c	—	7,348*	10,516*
BALB/c	GT	26,856**	44,080
BALB/c	GAT	5,788	10,152
GT			
—	DBA/2-APC+GT		2,070
+	DBA/2-APC		3,006
+	DBA/2-APC+GT		<u>35,450</u>
DBA/2-LPS blasts			
+	5 × 10 ⁵		2,768
+	2.5 × 10 ⁵		4,594
+	1.2 × 10 ⁵		4,484
+	6 × 10 ⁴		2,096
+	3 × 10 ⁴		1,554
+	1.5 × 10 ⁴		1,200
BALB/c-LPS blasts			
+	5 × 10 ⁵		<u>13,100</u>
+	2.5 × 10 ⁵		<u>15,040</u>
+	1.2 × 10 ⁵		<u>14,266</u>
+	6 × 10 ⁴		<u>11,726</u>
+	3 × 10 ⁴		<u>9,521</u>
+	1.5 × 10 ⁴		3,026

Top, priming with BALB/c LPS blasts. DBA/2 mice were injected intraperitoneally three times at weekly intervals with 10^7 irradiated (30 Gy) Ficoll separated blasts from BALB/c or DBA/2 spleen cells that had been cultured for three days with $10 \mu\text{g ml}^{-1}$ LPS. Seven days later spleen cells were tested *in vitro* for their responses to GT or GAT¹⁴ as described in Fig. 1. Bottom, priming with GT. LN cells from mice primed with GT (see Fig. 1) were cultured for four days with $150 \mu\text{g ml}^{-1}$ GT, washed and subsequently cultured for three days with 10% Concanavalin A (Con A) supernatant from rat spleen cells. * Residual proliferation to the injected BALB/c cells. ** Significant responses are underlined.

and, along with the dominance of unresponsiveness, hinted that the controlling gene might encode an antigen which mimics GT.

Where is the gene expressed?

To look for the site of expression, we created a set of chimaeric mice, whose bodies, thymuses and immune systems were made up of various combinations of R and NR cell types. Figure 2 shows the responses of 32 such chimaeras tested at the same time. Clearly, the response to GT is dependent on the presence of responder type haemopoietic cells and independent of the genotype of the thymus or the body of the animals. This result indicates that the gene controlling non-responsiveness is expressed only by the radiosensitive cells of the immune system, and not by any other tissues, including the radio-resistant antigen presenting cells (APCs) that remain throughout the thymus and body.

Non-responsiveness is due to self-tolerance

If unresponsiveness to GT in BALB/c mice is actually due to tolerance of a self-antigen expressed only by haemopoietic cells, then a small number of BALB/c cells should be sufficient to tolerize DBA/2 T cells and render them unresponsive to GT. To test this, we constructed chimaeras containing mixtures of DBA/2 and BALB/c fetal liver cells. We estimated the levels of chimaerism with a DBA/2 anti-BALB/c T-cell line. Figure 3a shows the proliferative response of this indicator line to increasing numbers of BALB/c stimulators. The response is linearly proportional and sensitive to very few BALB/c cells

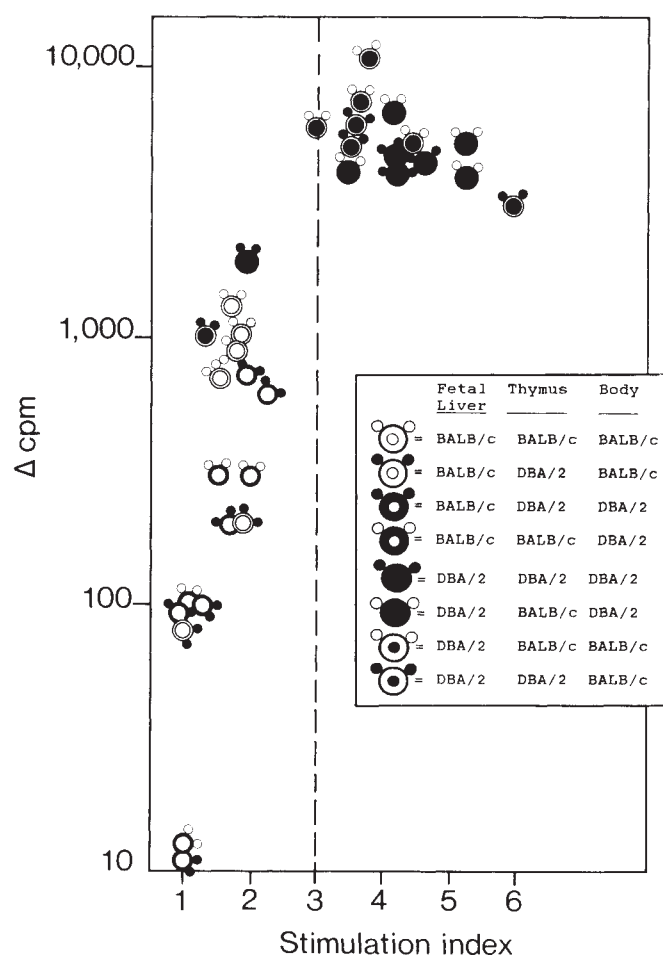


Fig. 2 Response to GT of mixed chimaeras is correlated only with the genotype of haemopoietic cells and not with that of the thymus or body. BALB/c and DBA/2 mice were thymectomized at five weeks of age, irradiated (9 Gy) at three months and immediately repopulated with 4×10^6 15–17-day-old BALB/c or DBA/2 fetal liver cells. Three weeks later each survivor was grafted under the left kidney capsule with an irradiated (6.5 Gy) thymus from BALB/c or DBA/2 mice that were less than 24 hours old. Four months later the animals were immunized with GT and tested as described in Fig. 1. The tissues of the chimaeras were mixed in all eight possible combinations of body, thymus and fetal liver (giving rise to haemopoietic cells) as in the key. Each point represents an individual mouse. BALB/c tissue is represented by open symbols, DBA/2 by closed symbols. Big circles, body; small circles, thymus lobes; inner circles, fetal liver.

(1.5%) in the stimulating population, allowing us to use it to determine low levels of BALB/c chimaerism.

Figure 3b shows the percentage of BALB/c cells present in each chimaera and its response to GT. Animals 3, 6 and 9 are the most informative: containing 3–6% BALB/c cells respectively, they were all unresponsive. This suggests that the unresponsiveness of BALB/c mice is indeed due to self-tolerance of an antigen expressed by a radiosensitive haemopoietic cell.

Molecular mimicry

If BALB/c cells express a molecular mimic of GT, then an injection of BALB/c spleen cells should immunize a DBA/2 responder mouse against GT and prepare it to make a secondary *in vitro* response. We found that spleen cells, thymocytes, dendritic cells and T-cell blasts immunize very poorly (data not shown). However, B cell blasts activated by lipopolysaccharide (LPS) immunized very well.

Table 1 (top) shows the responses of DBA/2 mice which were injected with LPS blasts. Unprimed mice and mice primed with

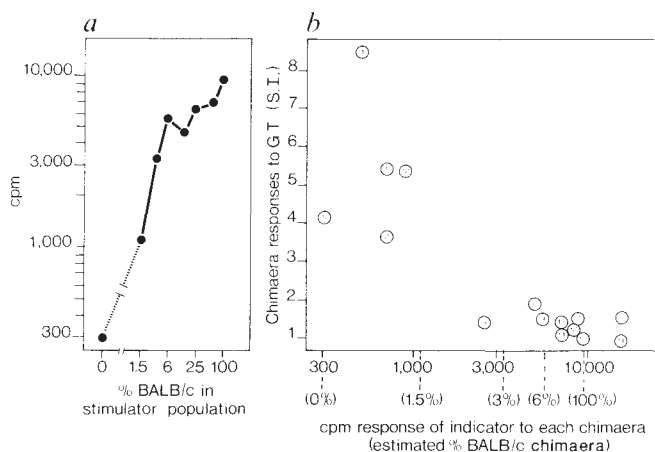


Fig. 3 Small numbers of BALB/c cells render a chimaera unresponsive. BALB/c and DBA/2 cells are difficult to distinguish. Although they differ by many cell surface molecules, very few of these can be recognized by antibodies. We therefore grew out a DBA/2 indicator T-cell line which responds specifically to BALB/c. *a*, Proliferative response of 1×10^5 cells from the DBA/2 anti-BALB/c indicator line against 3×10^5 irradiated (20 Gy) stimulator cells which consisted of a mixture of DBA/2 and BALB/c spleen cells varying from 0–100% BALB/c as indicated. *b*, In the same experiment, the chimaeras were tested for chimaerism and for their response to GT. Testing for chimaerism: 3×10^5 spleen cells from each chimaera were used to stimulate the indicator line. Each chimaera is numbered and placed on the X axis according to the response it induced in the indicator. The upper scale represents the indicator's response in c.p.m. In brackets is the estimated % BALB/c chimaerism, deduced from the indicator's response. The Y axis represents the responses of each chimaera to GT. Chimaeras 1–10 have BALB/c bodies, 11–15 have DBA/2 bodies.

DBA/2 blasts did not respond to GT, whereas mice primed with BALB/c LPS blasts responded to GT and not to the related antigen, GAT. Conversely, DBA/2 mice primed with GT responded to BALB/c LPS blasts but not to their own syngeneic blasts (Table 1, bottom). Thus the BALB/c LPS blasts express an antigen which mimics GT.

We have provisionally named the controlling gene 'Immune mimicking-gene for GT' or *Im^{gt}*. The capital *I* reflects the dominance of mimicry and unresponsiveness, and the superscript *gt* reflects its specificity.

Unlike GT, the product of *Im^{gt}* is not likely to be a random co-polymer of glutamic acid and tyrosine. Thus the two antigens are probably similar only in their MHC-associated forms.

To determine whether the *Im^{gt}* mimic was presented in a standard, MHC-restricted manner, we generated a T-cell hybridoma against GT (T-cell hybrid DGT 1) and compared its responses with that of lymph node (LN) populations from GT primed mice. Table 2 shows that DGT 1 behaves similarly to the LN populations. It responds to GT and BALB/c LPS blasts but not to DBA/2 blasts. This is direct evidence for molecular mimicry, showing that BALB/c LPS blasts express a mimic of GT + A^d. Responses to both antigens are blocked by antibodies against CD4 and against MHC class II molecules, showing that recognition of the two antigens, *Im^{gt}* and GT, requires the presence of the same cell membrane molecules: class II MHC as a restriction element on APCs and CD4 as an accessory molecule on T cells. There is also another system showing that B cells carry an antigen, not generally found on all leukocytes, which has an effect on T-cell responsiveness¹⁸. It will be interesting to see if the two antigens are related.

Having mapped a controlling gene in the B6 background¹⁴ we asked whether B6 cells expressed the same *Im^{gt}* product as BALB/c. Figure 4 shows that DGT 1 cells respond to BALB/c, (BALB/c × DBA/2)_F₁ and (B6 × DBA/2)_F₁ stimulators. Thus B6 also expresses a GT mimic and again mimicry, like unrespon-

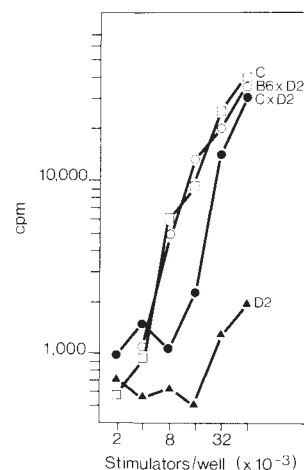


Fig. 4 Both the BALB/c and B6 genomes express *Im^{gt}*. The hybrid DGT1 was cultured for 24 h with spleen cells which had been activated with LPS and the supernatants were tested for activity as described in Table 2. C, BALB/c; D2, DBA/2; C × D2, (BALB/c × DBA/2)_F₁; B6 × D2, (B6 × DBA/2)_F₁.

siveness, is dominant. We do not know at present whether the *Im^{gt}* genes in B6 and BALB/c are identical or even allelic, but they do behave identically in three ways: (1) they confer dominant unresponsiveness to GT; (2) they produce (directly or indirectly) a determinant which is expressed on activated B cells and which mimics GT + A^d; and (3) the determinant is doubly dominant; it is expressed whether or not the gene itself or the MHC is in a heterozygous state.

Is *Im^{gt}* an oddity?

Because allelic variation is low in most non-MHC genes most cases of non-responsiveness would map entirely to the MHC. Many cases of mimicry may thus have gone unnoticed^{9,10}. However, there are a few cases where mimicry would seem to be the most likely explanation.

The most striking is the guinea pig response against the A chain of beef insulin¹². Strain 2 APCs cannot present beef insulin to their own T cells yet are perfectly able to stimulate T cells from strain 13. Clones of strain 13 T cells, which recognize MHC-2 + beef insulin, also respond to guinea pig insulin with MHC-2. The implication here is that the strain 2 MHC molecule combines with both beef and guinea pig insulin to form similar determinants and that strain 2 guinea pigs do not respond to the A chain of beef insulin because they are tolerant of their own mimic. Unfortunately, because of the MHC difference, the direct test of priming with the mimic is not possible. So this remains an inference.

Several other cases of dominant unresponsiveness seem to be maintained by suppression, a mechanism which has also been implicated in the maintenance of self-tolerance^{19–25}.

Self-mimicry may well be involved in the HLA-linked diseases in humans. In many cases the linkage to a particular MHC locus is not absolute and it is often postulated that the controlling gene is not the MHC itself but some other neighbouring gene. Perhaps these diseases are controlled by two genes, as is the response to GT: the MHC-restricting element and a non-MHC gene which has an indirect influence on the immune system. Diabetes is controlled this way in rodents^{26,27}, and careful family studies should determine whether it and other HLA-linked diseases show the same characteristics in humans.

What are the consequences of *Im* genes?

Sometimes the _F₁ hybrid between two responding strains will be unresponsive. For example if self 1 + MHC'A' mimics X + MHC'A' and self 2 + MHC'B' mimics X + MHC'B' then a strain carrying self 1 + B or self 2 + A will respond but the _F₁ hybrid

Table 2 Blocking responses to GT and Im^{gt} using monoclonal antibodies (mAb)

mAb (specificity)	Antigen (GT)	Responses to GT		Responses to Im^{gt}	
		LN population (c.p.m.)	DGT1 (c.p.m.)	Antigen (LPS blasts)	DGT1 (c.p.m.)
—	—	1,385	7,483	—	1,072
—	+	38,736	35,985	DBA/2	2,368
				BALB/c	21,125
8.A ($A\alpha\beta$)	+	35,013	33,132	NT	
39.I ($A\beta E\beta$)	+	6,172	4,852	BALB/c	1,708
82.B ($A\alpha\beta E\beta$)	+	5,689	3,265	BALB/c	4,113
40.I ($A\beta E\beta$)	+	25,641	25,294	NT	
10.A ($E\alpha$)	+	33,137	29,726	BALB/c	15,507
40.D ($E\beta$)	+	28,280	28,636	BALB/c	14,585
7.A ($E\alpha\beta$)	+	20,474	36,053	BALB/c	16,259
74.C ($E\alpha\beta$)	+	34,724	33,338	BALB/c	15,611
129.19.6.8 (CD4)	+	2,491*	5,472	BALB/c	1,536
H35-17.2 (CD8)	+	11,895*	33,275	NT	
H68-96.22 (LFA-1)	+	4,919*	3,839	NT	
—	—	1,092*			
—	+	15,035*			

The LN T-cell proliferation assay is described in the Fig. 1 legend. Hybridoma: DBA/2 mice were immunized with $100 \mu\text{g GT}^{14}$. Seven days later cells from the draining LN were cultured with $150 \mu\text{g ml}^{-1}$ GT. Surviving cells were fused with BW5147, a T lymphoma whose stimulation can be assayed by the production of the T-cell growth hormone, IL-2^{33,34}. In an attempt to raise clones which were as representative of the uncloned population as possible, we selected GT-specific hybridomas with $150 \mu\text{g ml}^{-1}$ GT, the dose which is most effective for the population. The resulting hybridoma DGT 1 (for DBA/2, GT specific) is Thy1.2⁺, CD4⁺, CD8⁺, specific for GT and not for GA or GAT. At the time these experiments were done the hybridoma had been re-cloned twice. Functional assay: 1×10^5 hybridoma cells were cultured with APCs (B-cell lymphoma M12.4.1, 3×10^4 per well) with or without GT. After 24 hours, the supernatants were harvested and tested at 75% concentration for their ability to support the growth of 10^4 CTL-L cells (an IL-2 dependent cell line) for 24 h. Data presented are the means of triplicate cultures. Antigens and mAbs: in every case the dose of GT was $150 \mu\text{g ml}^{-1}$ and of Ab was $5 \mu\text{g ml}^{-1}$, except that anti-CD4 was 100 ng ml^{-1} . All the mAb, directed against A and/or E were from M. Pierrés and are $\gamma 2a$. Anti-CD4 is $\gamma 2a$, anti-CD8 is $\gamma 2b$, and anti-LFA-1 is $\gamma 2b$ (refs 35–38). The mAbs were put in at the beginning of the 24-hour culture period. LPS blasts were obtained from 48-h spleen cell cultures with $10 \mu\text{g ml}^{-1}$ of LPS. The final dose of Ficoll-separated irradiated (30 Gy) blasts was 6×10^4 per well. * Data obtained from separate experiments.

will be unresponsive. Such a situation has been reported²⁸.

Responses to foreign antigens will occasionally lead to autoreactivity. An immune response against a foreign antigen could bring a self-mimic under relentless attack. Once started, such an immune response could be constantly boosted by the self-mimic. Perhaps this is why many HLA-linked diseases are autoimmune, and not a few of them occur following an immune response to a foreign antigen²⁹. Perhaps this is also why many cases of non-responsiveness involve suppression, a strong fail-safe mechanism³⁰.

There will be selection pressure against the expression of too many MHC genes. Because the number of cells in the immune system is finite, the elimination of cells recognizing 'self' obviously reduces the recognition repertoire of the individual. If for every new MHC product expressed, there must be tolerance of thousands of new combinations of MHC with other self-proteins, there will be a strong selection for individuals carrying a minimum load of expressed MHC loci. Opposing this will be the pressure to express as many MHC loci as possible, in case any one allele is inefficient at forming complexes with particular pathogens. The selection pressure for more MHC loci will be the weaker one for two reasons. Firstly an animal missing

an MHC allele necessary to fight a particular pathogen is in trouble only if that particular pathogen happens to exist in its neighbourhood and lifetime. Second, any potential pathogen will have numerous antigenic sites, only a few of which might mimic self. A balance between the two pressures can be achieved with a small set of loci per individual and a large number of variants in the population. An illustrative case is seen in *Xenopus* where polyploid species are polyploid for most expressed genes but diploid for the nucleolar organizer (for obvious reasons) and for the MHC^{31,32}.

Finally, it has been estimated that there are about 50,000 proteins in the body. If we must be tolerant of all their possible determinants in all of their various complexes with MHC, how do we have a functioning immune system left which can respond against the outside world?

We thank Dr Michel Pierrés for the monoclonal antibodies, Dr Benjamin A. Taylor for information on the Im^{gt} gene position and Drs Martin Flajnik, James Kaufman, Antonio Lanzavecchia, Charles W. B. Mackay and Charles Steinberg for critical reading of the manuscript. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche and Co. Ltd, Basel, Switzerland.

Received 18 July; accepted 30 September.

- Benacerraf, B. & McDevitt, H. O. *Science* **175**, 273–279 (1972).
- Schwartz, R. H. *Adv. Immun.* **38**, 31–201 (1986).
- Blomberg, B., Geckeler, W. R. & Weigert, M. *Science* **177**, 178–180 (1972).
- Cohn, M. et al. in *The Immune System, Genes, Receptors, Signals* (eds Sercarz, E. E., Williamson, A. R. & Fox, C. F.), 89–117 (Academic, New York, 1974).
- Epstein et al. *J. exp. Med.* **163**, 759–773 (1986).
- Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701–702 (1974).
- Babbitt, B., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359–361 (1985).
- Billingham, R. E., Brent, L. & Medawar, P. B. *Phil. Trans. R. Soc. Lond.* **239**, 357–414 (1956).
- Schwartz, R. H. *Scand. J. Immun.* **7**, 3–10 (1978).
- Matzinger, P. *Nature* **292**, 497–501 (1981).
- Clark, R. B. & Schevach, E. M. *J. exp. Med.* **155**, 635–640 (1982).
- Dos Reis, G. A. & Schevach, E. M. *J. exp. Med.* **157**, 1287–1299 (1983).
- Ishii, N., Nagy, Z. A. & Klein, J. *J. exp. Med.* **157**, 998–1005 (1983).
- Vidović, D., Klein, J. & Nagy, Z. A. *J. Immun.* **134**, 3563–3568 (1985).
- Taylor, B. A. in *Origin of Inbred Mice* (ed. Morse, H. C.) 423–438 (Academic, New York, 1978).
- Dembic, Z., Bannwarth, W., Taylor, B. A. & Steinmetz, M. *Nature* **314**, 271–273 (1985).
- Lee, N. E. et al. *J. exp. Med.* **160**, 905–913 (1984).

- Marrack, P. & Kappler, J. *Nature* **32**, 840–843 (1988).
- Urba, W. J. & Hildemann, W. H. *Immunogenetics* **6**, 433–445 (1978).
- Zaleski, M. B. *Immunogenetics* **3**, 226–238 (1974).
- Debré, P., Waltenbaugh, C., Dorf, M. & Benacerraf, B. *J. exp. Med.* **144**, 272–276 (1976).
- Debré, P., Kapp, J. A., Dorf, M. E. & Benacerraf, B. *J. exp. Med.* **142**, 1447–1454 (1975).
- Cinader, B., Koh, S. W. & Naylor, D. *Int. Archs Allergy appl. Immun.* **35**, 150–162 (1962).
- Dorsch, S. & Roser, B. *Nature* **258**, 233–235 (1975).
- Gorczynski, R. M. & MacRae, S. *J. Immun.* **122**, 747–752 (1979).
- Prud'homme, G. J. et al. *Immunol. Today* **6**, 160–162 (1985).
- Prochazka, M., Leiter, E. H., Serreze, D. V. & Coleman, D. L. *Science* **237**, 286–289 (1987).
- Gasser, D. L. & Shreffler, D. C. *Immunogenetics* **1**, 133–140 (1974).
- Botazzo, G. F. et al. *Immunol. Rev.* **94**, 137–169 (1986).
- Cohen, I. R. *Immunol. Rev.* **94**, 5–21 (1986).
- Du Pasquier, L. & Blomberg, B. *Immunogenetics* **15**, 251–260 (1982).
- Du Pasquier, L. et al. *Immunogenetics* **5**, 129–141 (1977).
- Köhler, G. & Milstein, C. *Nature* **256**, 495–497 (1975).
- Marrack, P. in *Isolation, Characterization and Utilisation of T Lymphocyte Clones* (eds Fathman, C. G. & Fitch, F. W.) 508–510 (Academic, New York, 1982).
- Pierrés, A. et al. *J. Immun.* **132**, 2775–2782 (1984).
- Pierrés, M., Devaux, C., Dosseto, M. & Marchetto, S. *Immunogenetics* **14**, 481–485 (1981).
- Pierrés, M., Goridis, C. & Golstein, P. *Eur. J. Immun.* **12**, 60–69 (1982).
- Naquet, P., Marchetto, S. & Pierrés, M. *Immunogenetics* **18**, 559–574 (1983).

Timing measurements of the binary millisecond pulsar in the globular cluster M4

J. McKenna & A. G. Lyne

University of Manchester, Nuffield Radio Astronomy Laboratories,
Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

Measurements of the periods of pulsars, and of their rate of change with time, are essential in understanding pulsar evolution. For binary pulsars, such as PSR1620–26 in the core of the globular cluster M4, orbital parameters can also be derived from timing measurements. Here we report the results of timing observations of PSR1620–26, including a first determination of its period derivative $\dot{P} = 3.2 \pm 0.4 \times 10^{-19} \text{ s s}^{-1}$ and its orbital eccentricity $e = 0.025316 \pm 0.000003$. The age of this millisecond pulsar is less than 2.2×10^8 years. The relatively high eccentricity is probably a result of gravitational encounters with other stars in the cluster during this time.

The binary millisecond pulsar PSR1620–26, discovered in the globular cluster M4 (NGC6121)¹, has been the subject of a timing programme at Jodrell Bank to measure its major spin and orbital parameters. Over a period of ~ 10 months, the pulsar has been observed at least once a week. Most of the observations have been made with the 76-m Lovell telescope but some of the early timing was undertaken with a nearby 30-m telescope. The measurements were made with de-dispersing receivers centred on a number of radio frequencies between 408 and 1,667 MHz. The timing analysis followed the usual procedures^{2,3}, the JPL planetary ephemeris DE200⁴ being used to convert pulse arrival times to the inertial frame of the Solar System barycentre. The pulsar's Very Large Array (VLA) position⁵, corrected for the elliptic terms of aberration, was assumed for this conversion. The results are summarized in Table 1 and include well determined values of the pulsar's spin period derivative and its orbital parameters.

The period derivative of PSR1620–26 indicates a surface magnetic field for the neutron star of 3.0×10^9 G (ref. 2) and a characteristic age ($P/2\dot{P}$) of 2.2×10^8 yr. The latter is an upper limit on the age of the millisecond pulsar.

Our measurements cannot directly determine the inclination, i , of the orbit to the line of sight. Assuming reasonable values of i , however, the orbital parameters of Table 1 indicate that if the mass of the pulsar is $\sim 1.4 M_\odot$, it has an orbital velocity of $\sim 10 \text{ km s}^{-1}$ and an orbital radius of $\sim 0.2 \text{ AU}$. The companion star is then $\sim 1 \text{ AU}$ distant from the pulsar and, with a probable mass of $\sim 0.3\text{--}0.4 M_\odot$, is almost certainly a white dwarf.

Millisecond pulsars are thought to be formed when a neutron star with a comparatively low magnetic field ($\sim 10^8\text{--}10^9$ G) accretes matter from a less massive companion star. This causes it to spin up towards a minimum period usually given by^{6,7}

$$P_{\min} \approx 2.4 B_9^{6/7} (M/M_\odot)^{-5/7} (\dot{M}/\dot{M}_{\text{Edd}})^{-3/7} R_6^{18/7} \text{ ms} \quad (1)$$

where B_9 is the magnetic field in units of 10^9 G, $\dot{M}_{\text{Edd}}$ is the Eddington accretion rate limit ($\sim 10^{-8} M_\odot \text{ yr}^{-1}$) and R_6 is the neutron star radius in units of 10^6 cm. During this phase, the system is visible as a low-mass X-ray binary. Upon completion of mass transfer, the neutron star may be observable as a radio millisecond pulsar, spinning down under the action of a magnetic braking torque. The concomitant period derivative determines the characteristic age and places an upper limit on the time elapsed since the end of the mass transfer stage, that is, on the age of the millisecond pulsar.

Figure 1 shows the $P\text{--}\dot{P}$ diagram for 360 pulsars, with PSR1620–26 and the other binary and millisecond pulsars identified. The 'spin-up line', representing the minimum spin-up

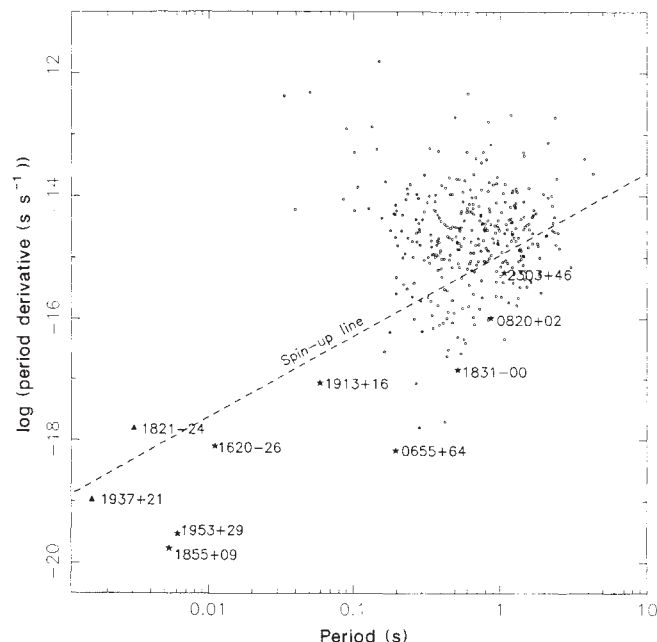


Fig. 1 The $P\text{--}\dot{P}$ diagram for 360 pulsars. The binary pulsars are marked with stars and the single millisecond pulsars with triangles. The conventional spin-up line, obtained using equation (1), is marked with a dashed line.

period and obtained using equation (1), is also shown. Because they subsequently spin down, millisecond pulsars should lie on or below this line. But the high period derivative of the millisecond pulsar in the globular cluster M28, PSR1821–24⁸, places it somewhat above the conventional spin-up line. White and Stella⁹ argued that previous analyses of the spin-up phenomenon had not taken into account the fact that radiation pressure dominates the inner accretion disk. They derived two alternative spin-up lines, both of which were situated above PSR1821–24 on the $P\text{--}\dot{P}$ diagram. PSR1620–26 lies below all of these lines.

After the mass transfer phase, the system consists of a pulsar in orbit about a white dwarf. The orbital eccentricities of such pulsars are often very small ($e \sim 10^{-4}$), the orbits having been made circular during the accretion process, but that of PSR1620–26 is much higher; PSR1620–26 has an eccentricity $e = 0.025$, which is, for example, two orders of magnitude larger than that of PSR1953+29 ($e = 0.00033$), an isolated binary pul-

Table 1 Measured parameters of the PSR1620–26 system

Pulsar period, P	$11075.7508061 \pm 0.0000002 \text{ } \mu\text{s}$
Epoch	47,200 modified Julian date (MJD)
Pulsar period derivative, $\dot{P}$	$(8.2 \pm 0.4) \times 10^{-19} \text{ s s}^{-1}$
Dispersion measure	$62.87 \pm 0.01 \text{ cm}^{-3} \text{ pc}$
Orbital period, P_b	$191.4427 \pm 0.0001 \text{ days}$
Projected semi-major axis, $a \sin i$	$64.8095 \pm 0.0001 \text{ light s}$
Eccentricity, e	0.025316 ± 0.000003
Longitude of periastron, ω	$117.12 \pm 0.01^\circ$
Epoch of periastron, T_0	$47196.719 \pm 0.004 \text{ MJD}$
Assumed position (DE200 coordinate system)	
Right ascension (1950.0)	$16^{\text{h}}20^{\text{m}}34.16 \pm 0.01^{\text{s}}$
Declination (1950.0)	$-26^{\circ}24'58.1 \pm 0.2''$
Derived parameters	
Mass function, $f(m_p, m_c)$	$0.008 M_\odot$
Pulsar surface magnetic field, B	$3.0 \times 10^9 \text{ G}$
Pulsar characteristic age, τ_c	$2.2 \times 10^8 \text{ yr}$

sar with an orbital period and radius similar to that of PSR1620–26¹⁰. We believe that the large eccentricity is probably the result of gravitational encounters with other stars in the globular cluster^{11,12}. Recent simulations¹³ predict that passing field stars in globular cluster M4 would take 3×10^8 yr to induce an eccentricity of 0.03 in the initially circular orbit of a pulsar such as PSR1620–26. This timescale is comparable to the pulsar's characteristic age. The binary pulsar systems that exist in regions of much lower stellar concentrations are not prone to such perturbations.

In the dense stellar environment of the globular cluster, the pulsar's low-mass companion is susceptible to ejection from the binary system in an exchange encounter with a more massive field star. However, the timescale for such exchange encounters is $\sim 6 \times 10^9$ yr (ref. 14). This is considerably greater than the characteristic age and supports the conclusion of Lyne *et al.*¹ that the pulsar's present companion was the mass donor that spun the pulsar up.

It has been pointed out¹⁵ that the movement of a millisecond pulsar through the stippled gravitational potential of a globular cluster core should affect the apparent spin parameters of the pulsar. In particular, we expect the pulsar's second period derivative to be considerably greater than predicted for magnetic

dipole braking. In the case of PSR1620–26, this should be measurable within a few years. The acceleration of the system's centre of mass should also result in an anomalous orbital period derivative of $\dot{P}_b \sim 10^{-11}$ s s⁻¹. Although this is an order of magnitude larger than that measured for PSR1913+16¹⁶, the long orbital period of PSR1620–26 means that such values will only be detectable after a few decades.

Received 6 September; accepted 7 October 1988.

1. Lyne, A. G., Biggs, J. D., Brinkow, A., Ashworth, M. & McKenna, J. *Nature* **332**, 45–47 (1988).
2. Manchester, R. N. & Taylor, J. H. in *Pulsars* (Freeman, San Francisco, 1977).
3. Backer, D. C. & Hellings, R. W. *Astr. Astrophys.* **24**, 537–575 (1986).
4. Standish, E. M. *Astr. Astrophys.* **114**, 297–302 (1982).
5. Goss, W. M., Kulkarni, S. R. & Lyne, A. G. *Nature* **332**, 47–49 (1988).
6. Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. *Nature* **300**, 728–730 (1982).
7. van den Heuvel, E. P. J. *Ann. New York Acad. Sci.* **302**, 14–35 (1977).
8. Foster, R. S., Backer, D. C., Taylor, J. H. & Goss, W. M. *Astrophys. J.* **326**, L13–L15 (1988).
9. White, N. F. & Stella, L. *Nature* **332**, 416–418 (1988).
10. Rawley, L. A., Taylor, J. H. & Davis, M. M. *Astrophys. J.* **326**, 947–953 (1988).
11. Hills, J. G. *Astr. J.* **80**, 809–825 (1975).
12. Heggie, D. C. *Mon. Not. R. astr. Soc.* **173**, 729–787 (1975).
13. Rappaport, S. A., Putney, A. & Verbunt, F. *Astrophys. J.* (in the press).
14. Hills, J. G. *Astr. J.* **82**, 626–630 (1977).
15. Blandford, R. D., Romani, R. W. & Applegate, J. H. *Mon. Not. R. astr. Soc.* **225**, 51P–53P (1987).
16. Weisberg, J. M. & Taylor, J. H. *Phys. Rev. Lett.* **52**, 1349–1350 (1984).

The ultraviolet extinction from interstellar graphitic onions

Edward L. Wright

Department of Astronomy, University of California, Los Angeles, California 90024-1562, USA

Carbon particles in interstellar dust cause absorption of starlight at certain wavelengths. Kroto and McKay¹ have suggested that such particles have the form of quasi-icosahedral spiral shells, with the hexagonal graphitic planes perpendicular to the radius, like the layers of an onion. Using the discrete dipole approximation (DDA)² I have calculated the ultraviolet extinction curves of small spherical particles with the tensor dielectric constant of graphite but with the *c*-axis always running parallel to the radius vector. The feature at a wavelength of 2,200 Å is doubled in width and moved to a wavelength about 80 Å shorter in comparison to the parameters derived from the usual Mie theory approximation for a mixture of particles with scalar dielectric constants. For larger particles, the peak shifts to longer wavelengths, but for a radius of ~ 300 Å which puts the peak at 2,200 Å (the same position as the mean interstellar absorption peak³), the width of absorption by quasi-icosahedral particles is too large to fit the data.

Graphite has an anisotropic dielectric constant, so the absorption efficiency of small graphite spheres cannot be calculated using the Mie theory. Draine (personal communication) has calculated the extinction from graphite crystals with the correct anisotropic dielectric constant, as illustrated in Fig. 1a, where the circles represent the discrete dipoles used in the calculation, and the line through the circle shows the orientation of the hexagonal planes in the graphite crystal. However, Kroto and McKay propose the onion-like structure shown in Fig. 1b for small carbon particles. I have used the discrete dipole approximation² (DDA) to compare the ultraviolet extinction of these 'Kroto-spheres' with the usual Mie model, which is a mixture of two types of non-existent particles: two thirds have an isotropic dielectric constant ϵ equal to $\epsilon(E \perp c)$ for graphite when the *E* field is perpendicular to the *c*-axis (and hence along the graphite planes), and the other one-third have an isotropic ϵ given by $\epsilon(E \parallel c)$ for graphite. Panels c and d in Fig. 1 illustrate these constructs. The tensor dielectric constant for graphite has

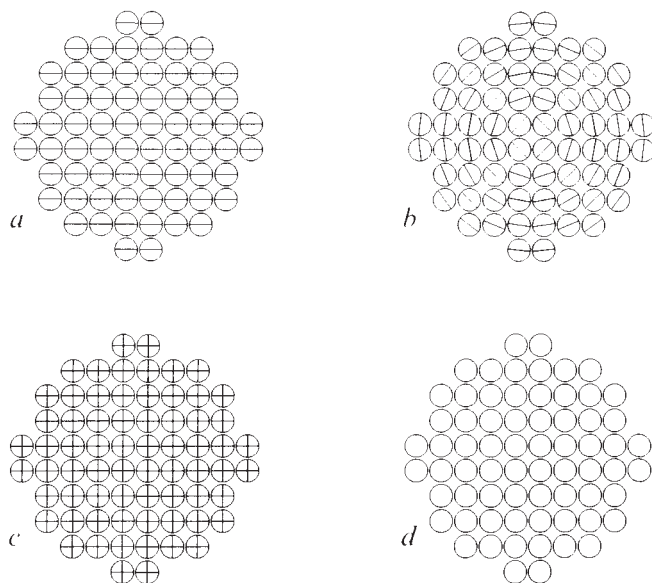


Fig. 1 Diagrams showing the arrangement of the graphite planes in the four models discussed in the text. a, Crystalline graphite; b, onion-like layers proposed by Kroto and McKay; for the Mie calculation I use a mixture in which two-thirds of the particles are like c and one-third are like d.

been taken as that used by Draine and Lee⁴. As I am primarily interested in the changes in the extinction curve caused by the onion-like geometry, I have computed both the onion geometry in Fig. 1b and the $\frac{2}{3}-\frac{1}{3}$ mixture using the DDA with exactly the same model geometry. Extinction curves for spherical models made from $N = 280$ discrete dipoles having radii $a = 50, 100$ and 200 Å are shown in Fig. 2. The solid curves are cubic splines through the calculated points for the Kroto-spheres; the dashed curves are fits for the $\frac{2}{3}-\frac{1}{3}$ mixture. The dots show the actual calculated values for the Kroto-spheres with $a = 100$ Å. The same set of frequencies was calculated for all cases. Draine's results (personal communication) show that randomly oriented spherical graphite crystals give extinction curves very similar to the dashed curves.

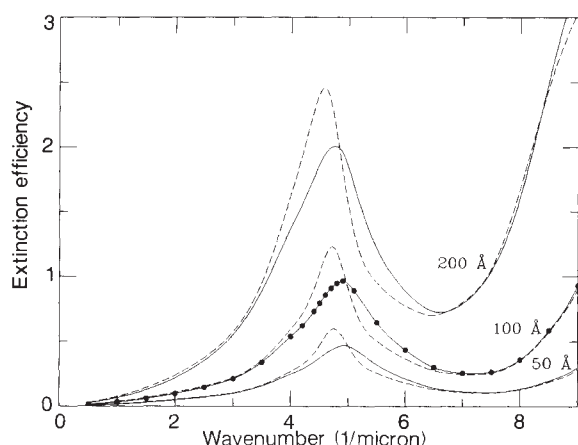


Fig. 2 Extinction efficiency $Q = \sigma/\pi a^2$ produced by the two models at three different radii. Solid curves are natural cubic splines through the points at which the DDA was computed for the kroto-spheres (these points are shown as dots on one curve), and the dashed curves are fits for the $\frac{2}{3}-\frac{1}{3}$ mixture. Total number of particles $N = 280$.

These curves of Q_{ext} versus λ have been compared to the mean observed interstellar extinction curve³ $A/E(B-V)$ using $A/E(B-V) = P_1 + P_2/\lambda + P_3 Q_{\text{ext}}(\lambda, a)$, where P_1 and P_2 describe the background continuous absorption, and P_3 adjusts the height of the 2,200-Å peak. The best fit for Kroto-spheres occurs for $a = 250$ Å, and for the $\frac{2}{3}-\frac{1}{3}$ mixture the best fit occurs for $a = 200$ Å. These fits are shown in Fig. 3, along with the fitted continuum. Figure 4 shows the results of a Drude model fit to the calculated curves, using the same form as that adopted by Fitzpatrick and Massa⁵ to analyse observed extinction curves. The rectangle shows the range of observed parameters towards different stars, whereas the curves show the behaviour of the graphitic onions and the $\frac{2}{3}-\frac{1}{3}$ mixture as a function of radius. Clearly the graphitic onions do not fit the 2,200-Å feature well, and cannot be a major constituent of the interstellar medium unless the curved geometry or chemical impurities such as hydrogen modify the dielectric tensor in a way that lengthens the peak wavelength and reduces the peak width for small particles.

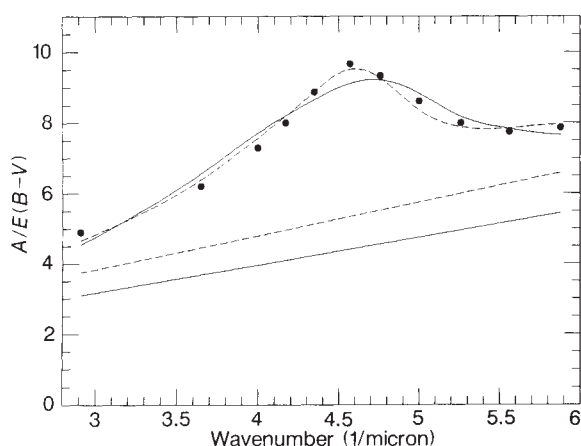


Fig. 3 Fit of calculated extinction (curves) plus a continuum (straight lines) to the observed mean interstellar extinction curve. The solid curve shows the best-fitting graphitic onions with $a = 250$ Å; the dashed curve shows the best-fitting $\frac{2}{3}-\frac{1}{3}$ mixture with $a = 200$ Å.

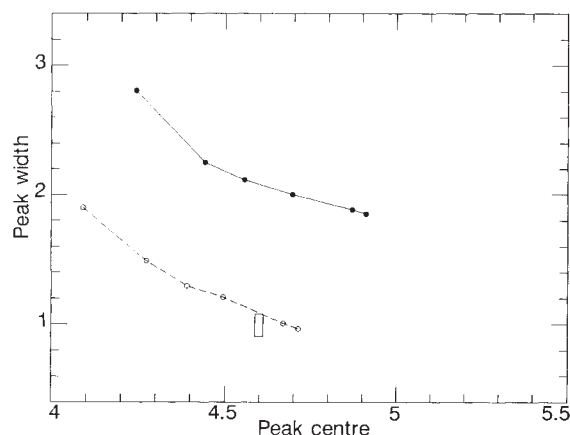


Fig. 4 Peak width and centre in μm^{-1} from a Drude model fit to observed data (rectangle) and to graphitic onions (solid curve, solid circles) and the $\frac{2}{3}-\frac{1}{3}$ mixture (dashed curve, open circles) as a function of radius. The points on the curves mark $a = 50, 100, 200, 250, 300$ and 400 Å from right to left.

Received 29 August; accepted 4 October 1988.

1. Kroto, H. W. & McKay, K. *Nature* **331**, 328-331 (1988).
2. Purcell, E. M. & Pennypacker, C. R. *Astrophys. J.* **186**, 705-714 (1973).
3. Savage, B. D. & Mathis, J. S. *A. Rev. Astr. Astrophys.* **17**, 73-111 (1979).
4. Draine, B. T. & Lee, H. M. *Astrophys. J.* **285**, 89-108 (1985).
5. Fitzpatrick, E. L. & Massa, D. *Astrophys. J.* **307**, 286-294 (1986).

An interstellar dust component rich in ^{12}C

R. D. Ash, J. W. Arden*, M. M. Grady, I. P. Wright & C. T. Pillinger

Planetary Sciences Unit, Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

* Department of Geology and Mineralogy, University of Oxford, Parks Road, Oxford OX1 3PR, UK

Chondritic meteorites contain grains which condensed in circumstellar environments and carry unusual isotope signatures¹ providing clues to the location in which grain formation took place. Isotope measurements on grains that are composed entirely of carbon or carbon compounds have, up to now, given compositions either close to typical Solar System values or enriched in ^{13}C (refs 2-5). Despite the existence of astrophysical environments where ^{12}C production predominates, highly ^{12}C -enriched carbon components in meteorites are uncommon. Here we report the results of an analysis of an acid-resistant residue from the Allende (CV3) meteorite^{1,6}, extensively combusted in oxygen at low temperatures ($<500^\circ\text{C}$) to remove isotopically normal carbon, which revealed the presence of a small amount (<0.1 p.p.m.) of carbon with a $^{12}\text{C}/^{13}\text{C}$ ratio close to 120, well above the value of 89 ± 3 found in terrestrial samples or bulk meteorites. This component is denoted CA. Stellar evolutionary processes favour lowering of the $^{12}\text{C}/^{13}\text{C}$ ratio; hence CA, which has a combustion temperature consistent with graphite, might be a primitive type of interstellar material.

Since the discovery of isotopically anomalous carbon (thought to be associated with interstellar dust) in carbonaceous chondrites¹, attention has been focused on CM2 meteorites, which contain abundant ^{13}C -enriched material²⁻⁵. The Allende CV3 meteorite has been studied less completely^{1,6}; its major carbonaceous interstellar dust component (C δ), now known to be

3-nm diamonds, is the carrier of both xenon enriched in heavy and light isotopes (XE(HL))⁷ and isotopically light nitrogen^{8,9}. The carbon isotopic composition of C δ ($\delta^{13}\text{C} = -38\%$; ref. 1) is close to that of the bulk carbon in some terrestrial organic materials but differs from that of the carbonaceous chondrites ($\sim -17\%$). Recent advances in isotope measurement techniques¹⁰ allow a much more detailed analysis of minor carbonaceous components in the Allende meteorite using stepped combustion¹¹. Isotopically normal^{1,12} amorphous carbon (C γ) constitutes most of the carbon in this meteorite; both this and the C δ component^{1,8} are readily combusted at temperatures below 550 °C. To look for minor carbonaceous components that are more stable to oxidation, an attempt was made to remove these major species by prolonged combustion at low temperatures, rather than by using harsh chemical oxidation procedures which are much less selective in removing carbon-bearing materials³.

Allende fragments were demineralized with 9M HF/1M HCl and 1M HCl, accompanied by intermittent agitation using ultrasound. The acid treatments were cycled until evaporation of the recovered acid yielded no visible residue. Sulphur (and possibly some isotopically normal carbon) was removed by oxidation in 16M HNO₃ warmed to $\sim 70^\circ\text{C}$ with an infrared lamp. After washing with water and acetone, and subsequent drying, the residue, called ARCIII (denoted CIII in ref. 13), was analysed and found to contain 38.2 wt% C with $\delta^{13}\text{C}$ of -16.2% (compared with whole rock values of 0.21 wt%, $\delta^{13}\text{C} = -17.2\%$, ref. 12). An aliquot of ARCIII was progressively burned in oxygen (obtained from the decomposition of copper oxide at 900 °C) at temperatures between 390 and 410 °C for periods of 52, 28, 28 and 76 hours. At regular intervals, carbon dioxide gas was removed for quantitative and isotopic analysis. When $\delta^{13}\text{C}$ of the CO₂ rose above $\sim -16\%$, precombustion was stopped and an aliquot of the residual ARCIII removed for detailed stepped combustion over the range from room temperature to 1300 °C. Progressive precombustion removed amorphous carbon (C γ) and diamond (C δ), and two ^{13}C -enriched carbon components were revealed which burned at temperatures of ~ 700 – 750 and 900 – $1,100^\circ\text{C}$; details of these are reported elsewhere¹³.

During the course of the above treatment, the total carbon content of ARCIII was lowered from the original 38.2 wt% to 3.3 wt%. To remove the remaining low-temperature carbon (over 90% of which was combustible at temperatures below 600 °C), the residue was precombusted for a further 52 hours at 500 °C to yield a new sample, Allende ARCIII(e), with a greatly diminished total carbon content (0.13 wt%). A stepped-combustion analysis of ARCIII(e) showed that, not only had residual low-temperature isotopically normal carbon been removed, but also the $\delta^{13}\text{C}$ signatures associated with the newly discovered ^{13}C -enriched components¹³ had declined, as a result of the presence of isotopically light material burning at the intermediate temperature of 750 °C. The light carbon may have previously been hidden in a mineral that was degraded by the precombustion treatment. Analysis of second aliquot of ARCIII(e), step-combusted in 25 °C steps over the period of interest (Fig. 1) indicated that the light components constitutes ~ 43 p.p.m. of the sample (corresponding to 0.06 p.p.m. in Allende whole rock) with a minimum $\delta^{13}\text{C}$ of $-261 \pm 2.4\%$, equivalent to a $^{12}\text{C}/^{13}\text{C}$ ratio of 120.4 ± 0.4 . This new component is designated CA.

A $\delta^{13}\text{C}$ value of -261% indicates that CA was not formed by normal Solar System processes, implying that we have identified, for the first time, isotopically light pre-solar carbon in a meteorite sample. Low $\delta^{13}\text{C}$ values have been encountered previously in bulk meteorites or in fractions isolated by chemical or physical processing, but they have either been sufficiently close to the terrestrial range for fractionation processes to provide a possible alternative mechanism for their production, or the errors involved in their measurement were so large that no significance could be attached. For example, a $\delta^{13}\text{C}$ of $-49.9 \pm 1.1\%$ has been measured for C θ , amorphous carbon that was released

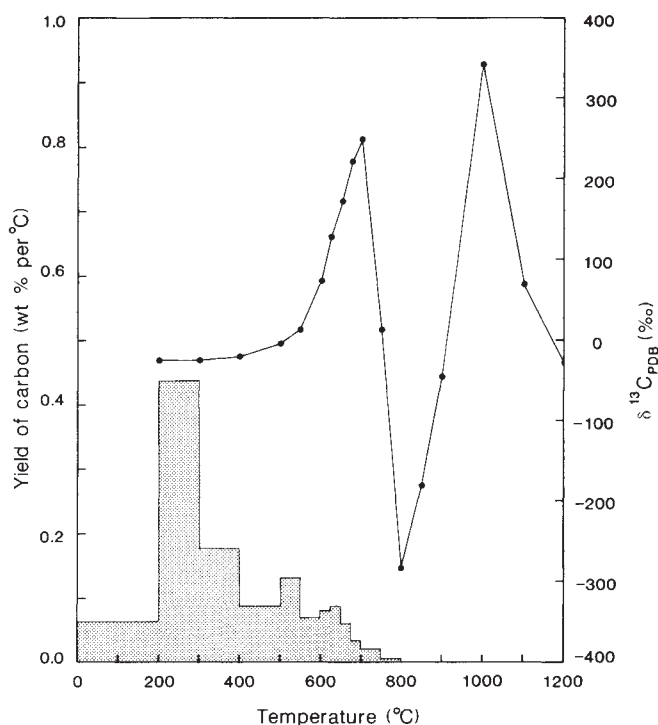


Fig. 1 Stepped combustion profile of Allende acid residue ARCIII(e). The histogram represents the carbon release (0.13 wt%) and the line provides isotope compositions at various temperatures, the minimum at 800 °C being -261% . ARCIII(e) was produced from ARCIII (a residue obtained by dissolving the Allende meteorite in HF/HCl; see text) by combusting at a constant oxygen pressure for 184 hours at between 390 and 410 °C followed by 52 hours at 500 °C. The relatively high abundance of carbon below 400 °C is a blank contribution introduced in the analytical system. PDB is the carbon isotope standard.

from Murchison spinel grains attacked with phosphoric acid³. Several ion probe studies have now been carried out on spinel grains, but mostly they show high $\delta^{13}\text{C}$ values^{14,15} which are thought to be associated with SiC (ref. 14), low values have also been encountered, but the errors of the measurements were $\sim \pm 46\%$ (refs 14, 16), limiting their significance, particularly as $\delta^{13}\text{C}$ studies of terrestrial spinel grains gave a spread of $0 \pm 150\%$ with the 1σ errors on individual measurements ranging from 10 to 140% (ref. 15).

Although previous measurements have not produced an unambiguous signature of isotopically light carbon, several pointers indicate that spinel is associated with ^{12}C -enriched carbon. C θ (amorphous carbon in spinel) was discovered only after attack of this mineral with phosphoric acid, but the presence of an isotopically light component had been suspected during earlier investigations (refs 3, 4, 17 and C.T.P., oral presentation at the Royal Society, London, 1986). It is known from studies on the Murray AY series of acid-resistant residues that oxidizing treatments, not noted for their ability to destroy or etch spinel, are able to render spinel grains porous to the oxygen used in stepped combustion³. It is equally possible that prolonged combustion at low temperatures (184 hours at 400 °C and 52 hours at 500 °C) degrades spinel sufficiently to allow oxygen access to carbon trapped within. Note that isotopically light carbon was recognized during ion-probe investigations of samples that had been exposed to plasma ashing^{14,16}. Spinel grains may be vulnerable under oxidizing conditions because they contain significant amounts of Fe²⁺; Palme and Wark¹⁸ report FeO contents of over 10% in spinel rims in both coarse and fine calcium aluminium inclusions (CAIs). Oxidation to Fe³⁺ may give rise to a network of faults within the grains,

causing degradation and disintegration. It may be significant that Zinner and Epstein¹⁴ found evidence for Fe₃O₄ and Fe₂O₃ in a sample that had been subjected to prolonged acid treatment before plasma ashing. It is possible that C λ is related to C θ , but it is in any case clear that either C λ is a form of carbon more stable to oxidation than is amorphous carbon, has a coarser grain-size, or is better protected by minerals than is C θ .

The nucleosynthetic production of ¹²C occurs through the triple α -process in stars undergoing helium burning. But in low-mass stars, before ejection in stellar winds or planetary nebula-forming events, ¹²C nuclei are subsequently processed through the CNO cycles to produce ¹³C, thus reducing the ¹²C/¹³C ratio. This ratio is lowered still further by CNO-processing of material that has been through more than one star-forming event. Thus the ¹²C/¹³C ratio is a sensitive indicator of continued cycling of matter by astration/star death. For example, type G, K and M red giant stars have ¹²C/¹³C ratios of 5–51 (ref. 19) whereas carbon stars have ¹²C/¹³C values of 30–70 (ref. 20) and the mean interstellar ratio is ~60–70 (ref. 21). Only in isolated cases do nucleosynthetic sites exhibit ¹²C/¹³C ratios that are elevated compared with solar values; for example, 95 $\pm$ 20 (from CN) or 100 $\pm$ 25 (from CO) have been observed in the carbon star U Cam²⁰, and a ratio in the range 40–130 has been obtained for CO in the well developed planetary nebular NGC 7027 (ref. 22). Clayton²³ suggests that the base of the He convection zone overlaying the core in massive stars might be a prolific source of ¹²C-rich dust. These observations notwithstanding, the tendency for ¹²C to be converted to ¹³C implies that the ratio of 120 measured in Allende is very unusual.

It has been proposed that ¹²C formed by core helium burning in massive stars will be ejected during supernovae events²⁴, but it is not clear whether the C/O ratio in such an environment is sufficiently high to condense elemental carbon⁵. If the C/O ratio became greater than 0.83 (a value in excess of the solar ratio of 0.6), carbon would be precipitated before spinel over the entire stability range²⁵. There are several astrophysical implications of the action of isotopically light carbon, formed under reducing conditions, as seed nuclei for spinel condensation³. Spinel grains concentrated by acid dissolution of Murchison^{26,27}, and those selected by hand from macroscopic high-temperature inclusions²⁸, seem to approximate to the ¹⁶O-rich end-member of the $\delta^{18}\text{O}$ – $\delta^{17}\text{O}$ relationship for CAIs, whereas those selected by freeze/thaw disaggregation and heavy liquid separation are at the other extreme in the trend¹⁶. The very smallest grains (0.2 μm) found in the Murray residue CF showed substantial deviations from the ¹⁶O mixing line²⁹. Both ¹⁶O and ¹²C are α -particle nuclei (that is, comprising four and three α -particles respectively), so if C λ were preferentially associated with ¹⁶O-rich spinels, then the isotope anomalies in C and O might have originated simultaneously in the same pre-solar supernova, possibly the one that precipitated formation of the Solar System³⁰. Carbon isotope measurements using stepped combustion directly on spinels and other minerals from CAIs (notably pyroxene and olivine, which show a ¹⁶O-rich signature) are necessary to test this suggestion. Some of these minerals are normally destroyed in acid dissolution experiments, but the existence of several minerals carrying light carbon is required to explain the presence of 'anti-heavy carbon' (C.T.P., oral presentation at the Royal Society, London, 1986) observed by McGarvie *et al.*⁴ in stepped combustion of HCl-resistant residues from CM2 and CI meteorites.

In searching for other possible formation sites for C λ , it is interesting to note that the few measurements available from comets³¹ imply ¹²C enrichment (¹²C/¹³C ~ 70–135) in these primitive bodies; these results, however, were unconfirmed by spacecraft encountering Halley's comet. Similarly, certain interstellar molecules and ions are predicted to be ¹²C-enriched³². Indeed, some interstellar molecules in the vicinity of the Solar

System have ¹²C/¹³C ratios in excess of the solar value^{33,34}. Such origins would imply an organic nature for C λ , which seems unlikely (see below). But in dense molecular clouds it is possible that the dust fraction may become ¹²C-enriched as a result of ion-molecule reactions of the type:



Assuming that in ARCI(e), C λ is no longer in a mineral-protected site, its combustion temperature may provide information about its nature. Stability up to 750 °C eliminates an organic nature, on the other hand, this temperature is too low for SiC (ref. 35). It would be compatible with diamonds of size > 60 μm (ref. 9) but interstellar diamonds in carbonaceous chondrites have a typical size of 3 nm, with a limit of 10 nm (ref. 36). Amorphous carbon tends to burn at temperatures below 700 °C; well-crystalline graphite combusts over the range 700–800 °C (ref. 37). Graphite has long been proposed as a stable stellar condensate at high C/O ratio and is predicted to occur as a major dust component in the interstellar medium on the basis of the feature at 220 nm in the starlight extinction curve³⁸. If C λ does turn out to be graphite then an omission from the catalogue of interstellar grains identified in primitive meteorites will be rectified³⁶. The ¹²C/¹³C ratios presented here would thus become significant, because such values cannot be obtained from interstellar dust spectroscopically. In contrast to the common isotopically anomalous carbon components isolated so far from carbonaceous chondrites, the high ¹²C/¹³C ratio of C λ indicates a limited contamination by the secondary CNO cycle, which might suggest its antiquity. Other authors independently are finding evidence of light carbon associated with SiC (ref. 39).

This work was supported by the SERC.

Received 19 July; accepted 27 September 1988.

- Swart, P. K., Grady, M. M., Pillinger, C. T., Lewis, R. S. & Anders, E. *Science* **220**, 406–410 (1983).
- Yang, J. & Epstein, S. *Nature* **311**, 544–547 (1984).
- Tang, M. *et al. Geochim. cosmochim. Acta* **52**, 1221–1234 (1988).
- McGarvie, D. W., Wright, I. P., Grady, M. M., Pillinger, C. T. & Gibson, E. K. *J. Mem. natn. Inst. Polar Res. spec. Iss.* **46**, 179–195 (1987).
- Zinner, E., Tang, M. & Anders, E. *Nature* **330**, 730–732 (1987).
- Ott, U., Mack, R. & Chang, S. *Geochim. cosmochim. Acta* **45**, 1751–1788 (1981).
- Lewis, R. S., Tang, M., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
- Lewis, R. S., Anders, E., Wright, I. P., Norris, S. J. & Pillinger, C. T. *Nature* **305**, 767–771 (1983).
- Ash, R. D. *et al. Meteoritics* **22**, 319 (1987).
- Carr, R. H., Wright, I. P., Joines, A. T. & Pillinger, C. T. *J. Phys. E. Sci. Instrum.* **19**, 798–808 (1986).
- Swart, P. K., Grady, M. M. & Pillinger, C. T. *Meteoritics* **18**, 137–154 (1983).
- Swart, P. K., Grady, M. M. & Pillinger, C. T. *Nature* **297**, 381–383 (1982).
- Ash, R. D. *et al. Lunar planet. Sci. XIX*, 15–16 (1988).
- Zinner, E. & Epstein, S. *Earth planet. Sci. Lett.* **84**, 359–368 (1987).
- Niederer, F. R., Eberhardt, P. & Geiss, J. *Meteoritics* **20**, 716–718 (1985).
- Grossman, L., Fahey, A. J. & Zinner, E. *Lunar planet. Sci. XIX*, 435–436 (1988).
- McGarvie, D. W., Grady, M. M., Gibson, E. K. Jr & Pillinger, C. T. *Meteoritics* **20**, 708 (1985).
- Palme, H. & Wark, D. A. *Lunar planet. Sci. XIX*, 879–898 (1988).
- Dearborn, D. S. P. in *CNO Isotopes in Astrophysics (IAU Symp. 67)* (ed. Audouze, J.) 39–44 (Reidel, Dordrecht, 1977).
- Lambert, D. L., Gustafsson, B., Eriksson, K. & Hinkle, K. H. *Astrophys. J.* **62**, 373–425 (1986).
- Wannier, P. G. A. *Rev. Astr. Astrophys.* **18**, 399 (1980).
- Knapp, G. R. & Chang, K. M. *Astrophys. J.* **293**, 281–287 (1985).
- Clayton, D. D. *Proc. Lunar planet. Sci.* **12B**, 1781–1801 (1981).
- Truran, J. W. in *CNO Isotopes in Astrophysics (IAU Symp. 67)* (ed. Audouze, J.) 145–154 (Reidel, Dordrecht, 1977).
- Larimer, J. W. & Bartholomay, M. *Geochim. cosmochim. Acta* **43**, 1455–1466 (1979).
- Clayton, R. N. & Mayeda, T. K. *Earth planet. Sci. Lett.* **67**, 157–161 (1984).
- McKeegan, K. D. *Science* **237**, 1468–1471 (1987).
- Clayton, R. N., Onuma, N., Grossman, L. & Mayeda, T. K. *Earth planet. Sci. Lett.* **34**, 209–224 (1977).
- Zinner, E. & Tang, M. *Lunar planet. Sci. XIX*, 1323–1324 (1988).
- Cameron, A. G. W. & Truran, J. W. *Icarus* **30**, 447–461 (1977).
- Vanysek, V. & Rahe, J. *Moon Planets* **18**, 441–446 (1978).
- Langer, W. D., Graedel, T. E., Frerking, M. A. & Armentrout, P. B., *Astrophys. J.* **277**, 581–604 (1984).
- Hawkins, I. & Jura, M. *Astrophys. J.* **317**, 926–950 (1987).
- Penzias, A. A. *Astrophys. J.* **273**, 195–201 (1983).
- Ash, R. D., Arden, J. W., Grady, M. M., Wright, I. P. & Pillinger, C. T. *Meteoritics* (in the press).
- Tang, M. & Anders, E. *Geochim. cosmochim. Acta* **52**, 1235–1244 (1988).
- Grady, M. M., Wright, I. P., Swart, P. K. & Pillinger, C. T. *Geochim. cosmochim. Acta* **49**, 903–915 (1985).
- Clayton, D. D. in *Chondrules and their origins* (ed. King, E. A.) (Lunar and Planetary Institute, Houston, 1983).
- Anders, E., Tang, M. & Zinner, E. in *Interstellar Dust (IAU Symp. 135)*, 9–10 (1988).

Large plasma velocities along the magnetic field line in the auroral zone

G. O. L. Jones*, P. J. S. Williams*, K. J. Winser†, M. Lockwood† & K. Suvanto‡

* Physics Department, University College of Wales, Aberystwyth SY23 3BZ, UK

† Rutherford-Appleton Laboratory, Chilton, Didcot OX11 0QX, UK

‡ Imperial College of Science and Technology, London SW7 2AZ, UK

In the auroral zone, ionospheric plasma often moves horizontally at more than 1 km s^{-1} , driven by magnetospheric electric fields, but it has usually been assumed that vertical velocities are much smaller. On occasions, however, plasma has been seen to move upwards along the magnetic field line at several hundred m s^{-1} . These upward velocities are associated with electric fields strong enough to heat the ion population and drive it into a non-thermal state^{1,2}. Here we report observations of substantial upwards acceleration of plasma, to velocities as high as 500 m s^{-1} . An initial upthrust was provided by a combined upwelling of the neutral atmosphere and ionosphere but the continued acceleration at greater heights is explained by a combination of enhanced plasma pressure and the 'hydrodynamic mirror force'³. This acceleration marks an important stage in the transport of plasma from the ionosphere into the magnetosphere.

In the night sector of the auroral ionosphere large upward velocities along the magnetic field line have occasionally been recorded by EISCAT, the European incoherent scatter radar⁴. These events have usually occurred several degrees of latitude to the north of EISCAT, and are associated with electric fields of 40 mV m^{-1} or more^{5,6}. The plasma velocity vector was deter-

mined by a matrix inversion of the three separate components of velocity as measured by the three EISCAT antennas. This gave an estimate of the field-aligned velocity at only a single height, which was subject to very large random errors from the unfavourable measuring geometry⁷. Large errors also affected the EISCAT measurement, even farther north, of downward velocities of 400 m s^{-1} (ref. 8).

On 6 May 1987, two such events were observed at 21:06 and 21:42 UT, farther south than usual and lying on the magnetic field line that passes through the EISCAT transmitter. As a consequence, the field-aligned plasma velocity $v_{\parallel}$ was measured directly at all heights between 150 and 550 km with errors of $\leq 30 \text{ m s}^{-1}$. The EISCAT measurements are summarized in Figs 1 and 2, which show variations with time and height respectively.

During the two events, $v_{\parallel}$ at 300 km lay in the range $50\text{--}90 \text{ m s}^{-1}$, but it increased steadily with height until at 550 km values of $300\text{--}500 \text{ m s}^{-1}$ were recorded. These upward velocities were equivalent to ion fluxes of $3\text{--}5 \times 10^{13} \text{ m}^{-2} \text{ s}^{-1}$, an order of magnitude larger than the usual steady-state ionospheric outflow (the 'classical polar wind'⁹). In general, we expect four factors to determine $v_{\parallel}$: (1) the meridional component of the neutral wind; (2) vertical upwelling of the neutral atmosphere as a result of strong, localized Joule heating; (3) the gradient of plasma pressure in the topside ionosphere; and (4) gravity. Under quiet conditions, (1), (2) and (4) combine to drive a low flux ($\sim 10^{12} \text{ m}^{-2} \text{ s}^{-1}$) of light ions from the top of the ionosphere. The application of a large electric field can, however, cause strong Joule heating of the neutral atmosphere in the upper E-region. It is estimated that under typical conditions the consequent upwelling of the atmosphere drives the plasma upwards with a velocity of $50\text{--}100 \text{ m s}^{-1}$, which is the upward velocity often seen in the midnight sector¹⁰. But these factors cannot explain velocities of 500 m s^{-1} , although such an initial upward velocity is essential in creating the opportunity for a much greater acceleration.

When the electric field is strong enough to drive the plasma perpendicular to the magnetic field line at a speed greater than the ion-acoustic speed, resonant charge exchange between the oxygen atoms and oxygen ions, and to a lesser extent other collisions, transfer energy to the ion population by frictional heating and change the ion velocity distribution from a Maxwellian towards a toroidal form^{1,2}. Once the ion velocity distribution becomes anisotropic, the plasma becomes subject to a 'hydrodynamic mirror force' as a result of the divergence of the magnetic field lines with increasing height³, and this acts in addition to the enhancement in upward pressure gradient from the increase in $T_{\parallel}$, the ion temperature measured along the magnetic field line (see Fig. 2a).

Near the F-region peak, O-O^+ collisions cause a frictional drag. If we accept the commonly quoted value of the O-O^+ collision cross-section this would inhibit any large increase in upward velocity. Recent measurements have indicated, however, that the collision frequency may be smaller by a factor of two to three¹¹ (although other reports suggest an increase¹²). In either case the frictional drag decreases rapidly with increasing height in proportion to the drop in oxygen density, so that if the upwelling of the neutral atmosphere can carry the ions above 400 km, thereafter the combination of enhanced plasma pressure and the mirror force will accelerate the plasma with decreasing resistance.

Using the lower value of $\nu_{\text{O-O}^+}$ together with the MSIS86 model of the neutral atmosphere¹³, we can estimate the frictional drag at each height. From the measured profiles of electron concentration, electron temperature and $T_{\parallel}$ we can determine the gradient of plasma pressure. From the measured strength of the electric field we can determine the mirror force. If we take as the boundary condition the observed upward velocity at 300 km (50 and 90 m s^{-1} at 21:06 and 21:42 UT respectively) we predict a height profile of upward plasma velocity that is in good agreement with the observations (Fig. 2b).

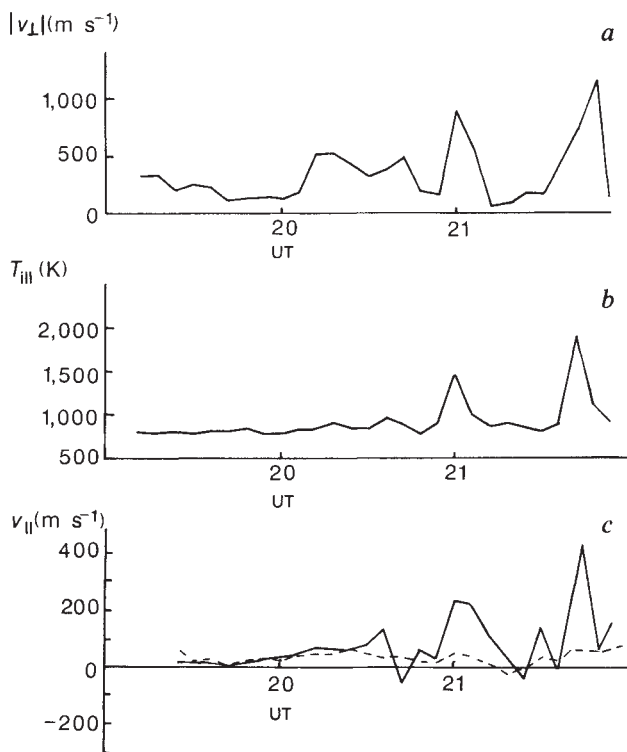


Fig. 1 Time variation of *a*, the magnitude of plasma velocity perpendicular to the magnetic field line; *b*, the ion temperature measured parallel to the field line at 300 km; and *c*, the plasma velocity parallel to the field line at 300 (dashed line) and 500 (solid line) km, as measured by EISCAT on 6 May 1987.

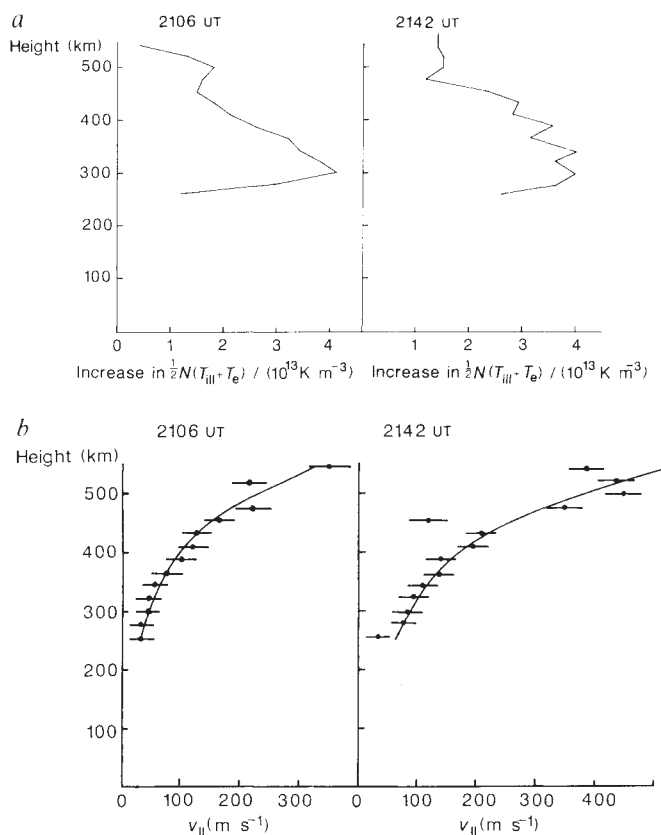


Fig. 2 Height profiles of *a*, the increase in plasma pressure at 21:06 and 21:42 UT on 6 May 1987 (*N* is the electron concentration and *T_e* is the electron temperature); *b*, the observed and model values of plasma velocity parallel to the magnetic field line during the two events. Observed values are dots with error bars, and predicted values are solid curves.

In determining the values of *T_{||}* we have assumed that in the topside of the ionosphere the ion population is dominated by *O⁺*. During upwelling it is possible that molecular ions might form a significant proportion of the ionospheric plasma at 300 km, in which case the true increase in *T_{||}* at this height would be even more than our measurements suggest. At greater heights we can be certain that the proportion of molecular ions is negligible, so that if upwelling of molecular ions occurs the effect of plasma pressure will be even stronger than we have calculated.

This observed acceleration is transient, not steady-state. A strong electric field obviously plays a key role, by driving the Pedersen currents which cause Joule heating and an upwelling of the neutral atmosphere. It also drives the ion population at high speed through the neutral atmosphere, which leads to frictional heating and an enhanced gradient of plasma pressure, as well as the formation of an anisotropic ion-velocity distribution which brings the hydrodynamic mirror force into play.

But a strong electric field is not enough in itself; if it were, large upward velocities would be more common. It is also necessary that soft particle precipitation (<1 keV) create sufficient conductivity in the upper E-region for the Joule heating to be intense. On 6 May 1987 such precipitation was indicated by the increased ionization at ~150 km which was observed between a sharp onset at 21:00 and an equally sharp cutoff at 21:48 UT. The degree of frictional heating is also much greater if the neutral velocity is in the opposite direction to the plasma velocity: these conditions occur in the region of the Harang discontinuity, and it is precisely in this region that the large upward velocities reported here were observed.

We acknowledge the constant support of the Director and

Staff at EISCAT. EISCAT is an international facility supported by France, Finland, West Germany, Norway, Sweden and the United Kingdom.

Received 8 August; accepted 10 October 1988.

1. St.-Maurice, J.-P. & Schunk, R. W. *Rev. Geophys. Sp. Phys.* **17**, 99-134 (1979).
2. Lockwood, M. *et al. Geophys. Res. Lett.* **14**, 111-114 (1987).
3. Suvanto, K., Lockwood, M. & Fuller-Rowell, T. G. *J. geophys. Res.* (in the press).
4. Folkestad, K., Hagfors, T. & Westerlund, S. *Radio Sci.* **18**, 867-879 (1983).
5. Winser, K. J., Jones, G. O. L. & Williams, P. J. S. *Physica Scripta* **37**, 640-644 (1988).
6. Winser, K. J., Jones, G. O. L. & Williams, P. J. S. *J. atmos. terr. Phys.* **50**, 379-382 (1988).
7. Jones, G. O. L., Winser, K. J. & Williams, P. J. S. *J. atmos. terr. Phys.* **48**, 887-892 (1986).
8. Smith, R. W., Winser, K. J., van Eyken, A. P., Quegan, S. & Allen, B. T. *J. atmos. terr. Phys.* **47**, 489-495 (1985).
9. Lockwood, M. *Adv. Sp. Res.* **6**, 63-77 (1982).
10. Winser, K. J., Jones, G. O. L. & Williams, P. J. S. *J. atmos. terr. Phys.* **48**, 893-904 (1986).
11. Winser, K. J., Farmer, A., Rees, D. & Aruliah, A. *J. atmos. terr. Phys.* **50**, 369-377 (1988).
12. Burnside, R. G., Tepley, C. A. & Wickwar, V. B. *Ann. Geophys.* **5A**, 343-349 (1987).
13. Hedin, A. *J. geophys. Res.* **92**, 4649-4662 (1987).

Where do channels begin?

David R. Montgomery & William E. Dietrich

Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA

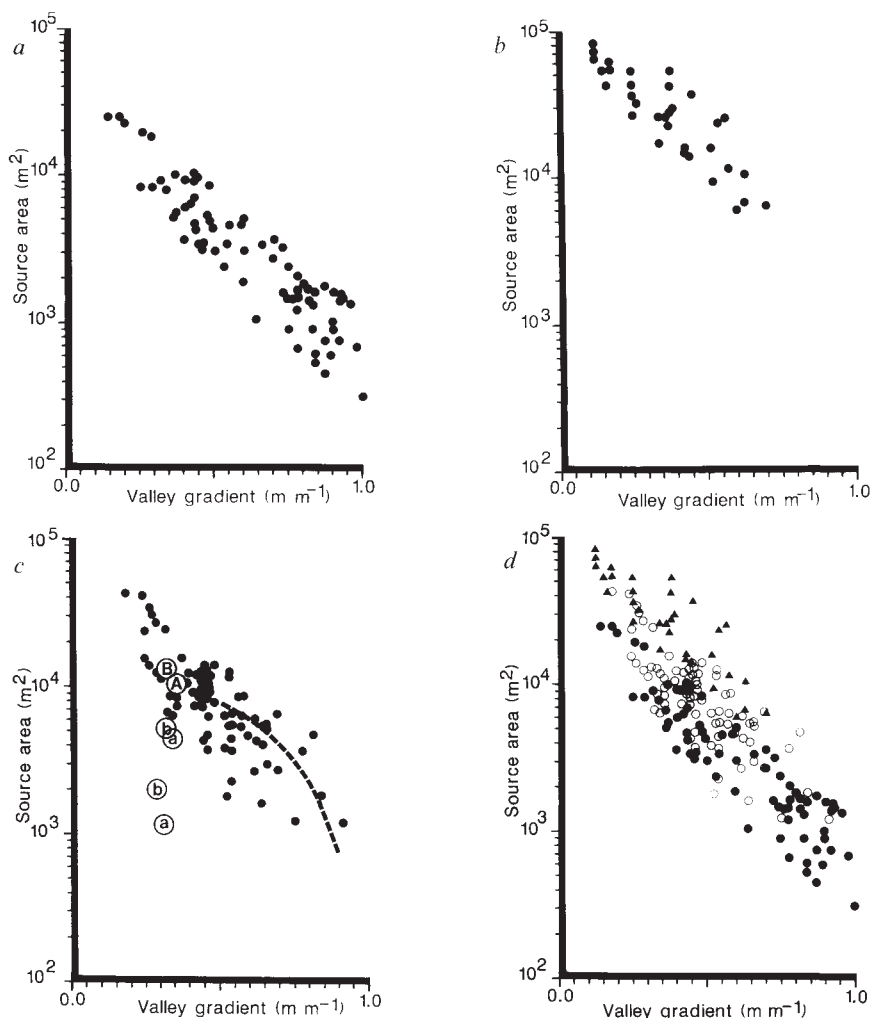
The closer channels begin to drainage divides, the greater will be the number of channels that occupy a unit area, and consequently the more finely dissected will be the landscape. Hence, a key component of channel network growth and landscape evolution theories¹⁻⁷, as well as models for topographically controlled catchment runoff⁸, should be the prediction of where channels begin. Little field data exist, however, either on channel head locations⁹⁻¹⁴ or on what processes act to initiate and maintain a channel¹⁴⁻¹⁷. Here we report observations from several soil-mantled regions of Oregon and California, which show that the source area above the channel head decreases with increasing local valley gradient for slopes ranging from 5 to 45 degrees. Our results support a predicted relationship between source area and slope for steep humid landscapes where channel initiation is by landsliding, but they contradict theoretical predictions for channel initiation by overland flow in gentle valleys. Our data also suggest that, for the same gradient, drier regions tend to have larger source areas.

We selected three areas in Oregon and California for a study of channel-head source areas on the grounds of accessibility, the range of slopes available, and their climatic and geological setting. The Coos Bay, Oregon, study area is located within an actively logged Douglas Fir (*Pseudotsuga menziesii*) forest on private land in the Oregon Coast Range, is underlain by folded Palaeocene basalts and gently dipping Eocene sandstones¹⁸, and receives a mean annual rainfall of ~1,500 mm (ref. 19). Of the 71 channel heads mapped in this area, 19 are associated with small-scale landslide scars, all of which predate recent logging activity. The plot of source area size against the local valley gradient at the channel head indicates a strong inverse correlation (Fig. 1*a*).

The southern Sierra Nevada study area is composed of two drainage basins on the Rankin Ranch, ~45 km east of Bakersfield, California. The region is covered by open oak woodland and grasslands, is underlain by Cretaceous granitic rocks²⁰ and receives an annual rainfall of ~260 mm (ref. 19). Evidence of previous landsliding and seepage erosion at abrupt channel heads was noted on steeper slopes; channels on gentle slopes, however, generally begin gradually and with evidence of overland flow. Most of the 33 channels mapped in this area begin as discontinuous channel segments. We defined a channel head as the farthest upslope location of a channel with well defined banks. The plot of source area against local hillslope gradient at the channel head for the southern Sierra data (Fig. 1*b*) also defines a clear inverse relationship. These data differ

Fig. 1 Source area plotted against the local valley gradient, or slope at channel heads, $\tan \theta$, for *a*, Coos Bay, Oregon, *b*, southern Sierra Nevada, California, and *c*, Marin County, California. For the study area in *c*, downslope from the drainage divide in two hollows the slope within the source area is relatively constant (lower case letters in open circles) until the channel begins (capital letters in open circles). The dashed line is the relationship predicted by equation (1). *d*, Combined data from all three field areas, defining a strong inverse linear relation over the full range of data. The southern Sierra data (triangles) generally overlie the Marin County data (open circles) which, in turn, appear to overlie the Coos Bay data (solid circles). This difference is most clearly expressed in the low gradient sites.

Methods. *a*, Channel head surveys were conducted at two recent clear cuts, the larger of which contained drainage basins composed of both low- and high-gradient source areas. The smaller basin consisted of moderately steep source areas in a single drainage basin. Every channel head in both areas was located in the field and mapped onto a 1:4,800 scale topographic basemap. The local hillslope gradient at each channel head was measured by two observers with Brunton compasses to within one degree agreement. Source areas contributing runoff to channel heads were measured with a digital planimeter for each of the three field areas. *b*, Channel head surveys were conducted in two adjacent basins, one of which contained predominantly shallow slopes whereas the other consisted primarily of steeper gradients. Field work used methods identical to those employed in the Coos Bay study area and was mapped onto an enlarged copy of the USGS Oiler Peak 7.5-min quadrangle. *c*, Source areas from three contiguous basins were mapped, using methods identical to those employed in the other study areas, onto an enlarged copy of the USGS Point Bonita 7.5-min quadrangle. Data for the unchannelled areas above channel heads were collected using identical methods. The procedure used to compare equation (1) with field data is discussed in detail in ref. 14 and by D.R.M. and W.E.D. (manuscript in preparation).



from those previously reported for this area¹³ in that the gradients presented here are field measurements of the unchannelled valley floor immediately above channel heads rather than the average slope of the entire source area measured from topographic maps. This distinction is important in that erosional processes controlling channel head locations should be strongly influenced by the local slope. Furthermore, elevations on published topographic maps are only accurate to within one half of the contour interval²¹, which would tend to artificially scatter map-derived data.

The Marin County study area occupies three contiguous drainage basins in the Golden Gate National Recreation Area 10 km north of San Francisco, California. In contrast to the relatively undeformed bedrock of the Coos Bay and southern Sierra areas, this region is underlain by stacked thrust sheets of greenstone, greywacke and chert typical of the Jurassic/Cretaceous Franciscan assemblage in the Marin Headlands terrane²². This area receives an average of 760 mm of rainfall annually²³ and was grazed before the early 1970s. Vegetation presently consists of coastal prairie, northern coastal scrub and introduced species. Of the 80 first-order channels mapped in this area, 49 begin abruptly, generally at small-scale landslide scarps or arcuate tension cracks. The remaining 31 channels begin as discontinuous channel segments with evidence for headcut undermining by seepage erosion. In both cases we defined the channel head location as the point nearest the upslope drainage divide that exhibits evidence of a channelled morphology. The plot of source area against local hillslope

gradient at the channel head for the Marin County data (Fig. 1c) also indicates an inverse relationship.

To test whether the inverse source-area/slope relation is a reflection of the well-known observation that the area of drainage basins increases with decreasing gradient downslope^{9,24}, we determined the area/slope relationship from the drainage divide downslope to the channel head, for two typical Marin County sites (Fig. 1c). The local valley gradient within these source areas is fairly uniform and the drainage area for locations along the topographic axis of both source areas fall below the variance of the data for channel heads. Furthermore, in both cases the channels begin at a point for which the drainage area lies within the scatter of the area/slope relationship for channel heads. These observations suggest that, in general, channels begin at the first point downslope from the drainage divide for which there is sufficient area to support a channel, and that the area/slope relation for drainage basins does not apply upslope of channel heads.

Two quantitative channel initiation theories^{14,25} for semi-arid to humid landscapes have been proposed which predict that source area size should vary as a function of valley gradient. For steep terrain, both theories assume that channel heads are formed by periodic, small-scale landsliding and both predict a rapid decrease in source area size with increasing valley gradient. Because of its analytical form and the limited number of variables it involves, the model by Dietrich *et al.*¹⁴ can be compared with our field results. The model is composed of two parts: a model for shallow sub-surface runoff (based on a theory by

Iida²⁶) and an infinite-slope, Coulomb failure model which, in this case, assumes that the soil is cohesionless at failure. Sub-surface runoff increases pore pressures in the soil, reducing the effective normal stress and consequently the strength of the soil. The two models can be combined by means of the predicted saturation depth (z), to relate the contributing drainage area (A), assumed to be the entire source area, to the local valley gradient ($\tan \theta$) and the angle of internal friction of the soil (ϕ):

$$A = (\rho_s / \rho_w) (K / R_0) z W \sin^2 \theta [(1 / \tan \theta) - (1 / \tan \phi)] \quad (1)$$

where ρ_s , ρ_w , K , W and R_0 are, respectively, the saturated bulk density of the soil, the fluid density, saturated hydraulic conductivity, width of the sub-surface runoff zone at the channel head and the steady-state rainfall generating the runoff. Equation (1) is valid only if pressures are hydrostatic and if the slope is less steep than the angle of internal friction, that is, when $[(\rho_s - \rho_w) / \rho_s] \tan \phi < \tan \theta \leq \tan \phi$.

All of these parameters except R_0 can be estimated from field observations and laboratory tests; at present R_0 must be a best-fit parameter in the model (D.R.M. and W.E.D., manuscript in preparation). For the Marin County study area the fitted R_0 has the value 8.5 mm h^{-1} , and equation (1) predicts a strong decrease in source area with increasing gradient (Fig. 1c), similar to that observed in the field data. Equation (1) also implies that a large variance in the source area gradient should be expected because the strength and saturated conductivity of the soil may vary considerably between sites. Because the channel head is displaced downslope when landslide scars infill, the time since the last landslide event at a site will also contribute to this variance.

Contrary to the inverse relationship observed for the data from steep slope gradients, one of the channel initiation theories predicts that in gentle valleys, where channel initiation is due to saturation overland flow²⁵, source areas will be positively correlated with valley gradient. This discrepancy suggests that further work is needed to refine channel initiation theories for shallow-gradient slopes, for which landsliding is uncommon.

Collectively, our data define a strong inverse linear trend over their full range (Fig. 1d). Within this general trend, however, the southern Sierra Nevada data generally plot above data from the Marin County study area, which, in turn, plot above the Coos Bay data. Although many other differences exist between the study areas, this sequence correlates with the variation in mean annual precipitation: 260 mm in the southern Sierra, 760 mm in Marin County and 1,500 mm in the Coos Bay area. Because runoff must increase with increasing drainage area, we suggest that for a given local slope, the source area size required to initiate a channel should increase with increasing aridity to produce the same critical combination of runoff and local slope at the channel head. This hypothesis has important implications for interpreting the response of low-order streams to climatic change. Smaller source areas in wetter regions suggest that channel heads would advance downslope in response to increased aridity, resulting the aggradational infilling of the upper reaches of former first-order channels²⁷. Similarly, channel heads would be expected to advance upslope into colluvial deposits during wetter periods.

Although most channel network evolution models assume that network growth occurs by branching and extension of network tips, the observation that many of the hillside channels in both the Tennessee Valley and southern Sierra areas are discontinuous suggests that channel head locations in these areas are generally controlled by hillslope processes. Furthermore, the inverse relationship of the field data indicates that drainage density should increase significantly with mean landscape gradient. In essence, the smooth, long hills typical of low relief areas result from the large source area required to initiate a channel, whereas the dissected terrain typical of steeper areas reflects the smaller source areas required to sustain channel processes.

We thank Preston Jordan, Anne Bikié, Kent Rich and Yutaka

Komatsu for providing field assistance and Steve Reneau, Peter Whiting, Tom Dunne and Jim Kirchner for constructive comments. We also thank Bill Rankin, the Weyerhaeuser Corporation and the Golden Gate National Recreation Area for access for our study areas. This research was supported in part by a NSF grant and matching funds from the Weyerhaeuser Corporation.

Received 6 June; accepted 5 October 1988.

1. Leopold, L. B. & Langbein, W. B. *U.S. Geol. Surv. Prof. Paper 500A* (1962).
2. Howard, A. D. *Geogr. Anal.* **3**, 29–50 (1971).
3. Abrahams, A. D. *Geol. Soc. Am. Bull.* **83**, 1523–1530 (1972).
4. Kirkby, M. J. in *Thresholds in Geomorphology* (eds Coates, D. R. & Vitek, J. D.) 53–73 (Allen & Unwin, Boston, 1980).
5. Kochel, R. C., Howard, A. D. & McLean, C. in *Models in Geomorphology* (ed. Woldenberg, M. J.) 313–341 (Allen & Unwin, Boston, 1985).
6. Dunne, T. & Aubry, B. in *Hillslope Processes* (ed. Abrahams, A. D.) 31–53 (Allen & Unwin, Boston, 1986).
7. Willgoose, G. R., Bras, R. L. & Rodriguez-Iturbe, I. *EOS* **69**, 345 (1988).
8. Beven, K. J. & Kirkby, M. J. *Hydrol. Sci. Bull.* **24**(1), 43–69 (1979).
9. Leopold, L. B. & Miller, J. P. *U.S. Geol. Surv. Prof. Paper 282-A* (1956).
10. Morisawa, M. *EOS* **38**(1), 86–88 (1957).
11. Maxwell, J. C. *Tech. Rep. 19, Off. Nav. Res. Proj. 389-042* (1960).
12. Mark, D. M. A. *Ass. Am. Geogr.* **73**, 358–372 (1983).
13. Dietrich, W. E., Reneau, S. L. & Wilson, C. J. *Proc. Int. Symp. Erosion Sedimentation Pacific Rim*, publ. no. 165, 27–37 (Int. Ass. Hydrol. Sci., 1987).
14. Dietrich, W. E., Wilson, C. J. & Reneau, S. L. in *Hillslope Processes* (ed. Abrahams, A. D.) 361–388 (Allen & Unwin, Boston, 1986).
15. Horton, R. E. *EOS* **13**, 350–361 (1932).
16. Kirkby, M. J. & Chorley, R. J. *Bull. Int. Ass. Sci. Hydrol.* **12**, 5–21 (1967).
17. Dunne, T. & Black, R. D. *Wat. Resour. Res.* **6**, 1296–1311 (1970).
18. Beaulieu, J. D. & Hughes, P. N. *Bull. Oregon Dept. Geol. Min. Ind.* **87** (1975).
19. Ruffner, T. A. *Climates of the States* (Gale Research Co., Detroit, 1985).
20. Calif. Div. Min. & Geol. *Geologic Map of California Bakersfield Sheet* (1964).
21. U.S. Geol. Surv. *Topographical Instructions* **1B4** (1952).
22. Wahrhaftig, C. in *Franciscan Geology of Northern California* (ed. Blake, M. C.) 31–50 (Soc. Econ. Paleontol. & Min., Los Angeles, 1984).
23. Rantz, S. E. *U.S. Geol. Surv. HA 298* (1968).
24. Hack, J. T. *U.S. Geol. Surv. Prof. Paper 294B* (1957).
25. Kirkby, M. J. in *Geomorphological Models* (ed. Ahnert, F.) 1–14 (Cremlingen, FRG, 1987).
26. Iida, T. *Trans. Jap. Geom. Un.* **5**, 1–12 (1984).
27. Reneau, S. L., Dietrich, W. E., Dorn, R. I., Berger, C. R. & Rubin, M. *Geology* **14**, 655–658 (1986).

Asthenospheric viscosity inferred from correlated land–sea earthquakes in north-east Japan

Paul A. Rydelek & I. Selwyn Sacks

Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, Washington, DC 20015, USA

The viscosity of the Earth's mantle has been estimated from studies of post-glacial rebound^{1,2}, post-seismic deformations of the ground following large earthquakes^{3,4}, and aftershock sequences^{5–7}. Here we derive a value for the viscosity of the asthenosphere from a correlation found in the historical catalogue of subduction-induced seismicity between the intraplate (land) and interplate (sea) earthquakes in north-east Japan. The correlation persists since the time of reliably reported earthquakes in AD 1600; land events precede sea events by ~36 yr, with a mean distance between land–sea pairs of ~200 km. Because of the viscoelastic coupling of the lithosphere to the asthenosphere, a plausible mechanism to explain the correlation is stress migration, governed by the viscosity of the asthenosphere. Large land shocks generate diffuse-like stress pulses which sweep past, and unlock, the thrust fault in the subduction zone, thus triggering the sea events. The correlation time and distance provide a measure of the speed of diffusion (5.6 km yr^{-1}) and hence an estimate of the viscosity ($7 \times 10^{18} \text{ Pa s}$).

The historical catalogue was obtained from a map by the Japanese Meteorological Agency (JMA), which lists the dates, magnitudes, and locations of events of Richter magnitude $M \geq 6.4$ reported since the seventh century AD. The JMA map is shown in Fig. 1 along with major subduction zones in the vicinity of Japan; these subduction zones are associated with active seismic regions, as is generally the case elsewhere. The reporting of earthquakes in the distant past, and the location and intensity

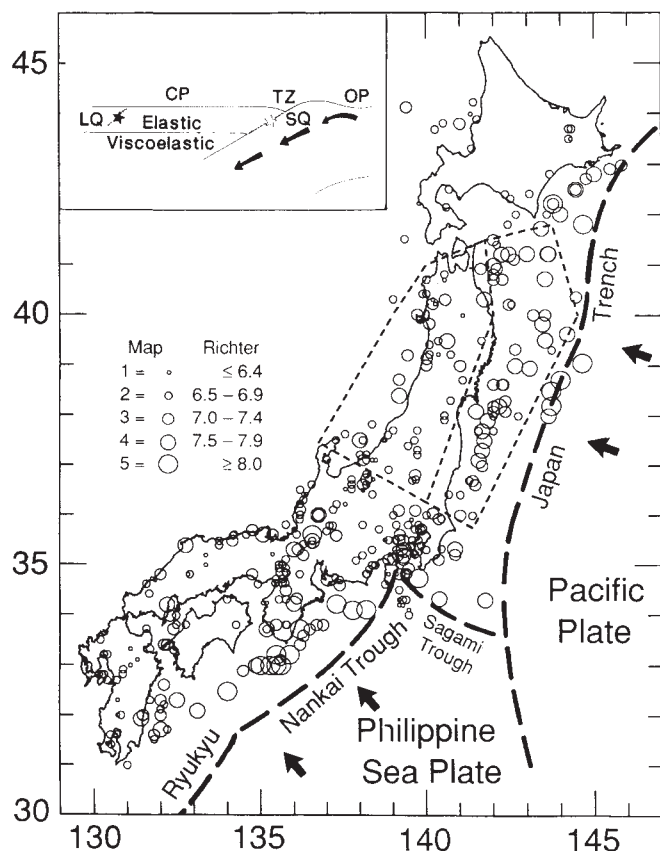


Fig. 1 The Islands of Japan with the locations and magnitudes of large historical and modern earthquakes compiled by the JMA. Conversion from map magnitude (1–5) to Richter magnitude is given in the legend. The dashed boxes in Tohoku, north-east Japan, are used to separate the land (intraplate) quakes from the sea (interplate, thrust zone only) quakes used in this study. The inset shows the collision between the continental (CP) and oceanic (OP) plates. The compression of the land plate from the sea plate in the thrust zone (TZ) will eventually produce a large land earthquake (LQ). The post-seismic decompression will diffuse outward and may trigger a sea earthquake (SQ) in the locked thrust zone.

assessment, was dependent on the population density, but after AD 1600 (the time of the unification of Japan) the catalogue is relatively complete with regard to the larger events ($M > 6.5$).

We isolated a section of north-east Japan (Fig. 1) because here the subduction zone of the Pacific plate has several unique features: (1) The zone is well isolated from adjacent subduction zones; (2) The catalogue provides a reasonably adequate statistical sampling of seismicity both on land and in the thrust zone (sea events); and (most importantly) (3) the motion of the subducting slab, inferred from studies⁸ of deep and intermediate focus earthquakes, fits a simple model of a dense sinking slab. In this region, the old (>100 Myr) Pacific plate subducts at approximately normal incidence to the Japan coast, and once past the thrust zone the slab sinks with minimal contortion at about 30° dip to deeper than 400 km.

A plot of the time history (1600–1971) of the land and sea earthquakes in the isolated section is shown in Fig. 2a. The sudden increase in the rate of reported quakes (especially for the sea events) after 1890 is a direct result of the establishment of the Imperial Meteorological Agency (now the JMA) and modern seismic instrumentation. Because of this change, three time windows can be identified: window A is from 1600–1890 with 24 land and 14 sea events, window B is from 1890–1971 with 22 land and 36 sea events, and window C is the sum of A and B. Thus windows A and B form independent catalogues.

The land–sea earthquake records were transformed into a continuous time series by interpolation with zeros, and then

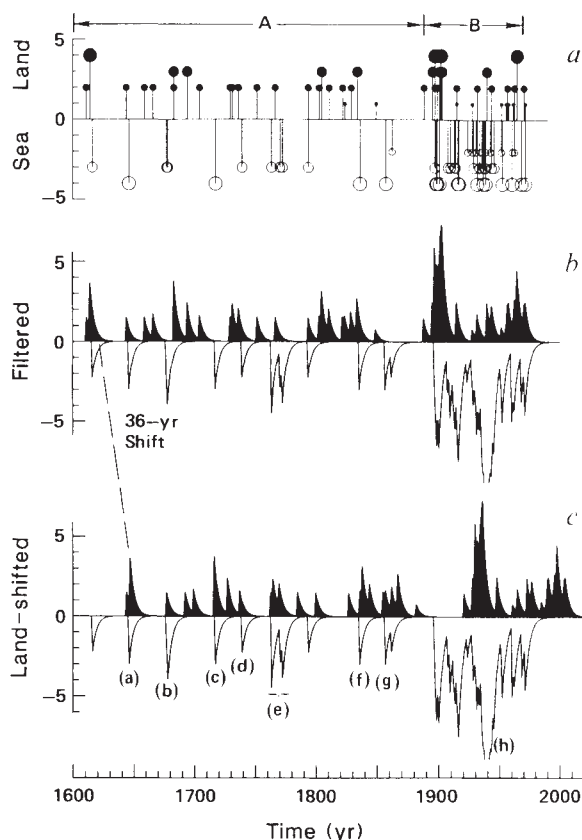


Fig. 2 Records of the land and sea events from the boxed sections of Fig. 1: *a*, Original data in which circles and sticks (solid for land events) are scaled to map magnitudes. *b*, High-pass filter with a time constant of 3 yr applied to the data. *c*, Land record time-shifted by 36 years. Windows A and B divide the catalogue into independent records with mostly stationary rates of earthquake occurrence; window C is A + B.

smoothed with a numerical high-pass single-pole filter. A filter time constant of 3 yr was chosen as this seemed to group together most of the obvious clusters of events. Thus earthquakes occurring within 3 yr appear as lengthy drawn-out events, whereas earthquakes or clusters of earthquakes separated by longer times appear isolated. The results using this filter were found to be stable to changes up to $\pm 50\%$ in the time constant. The filtered land–sea earthquake records shown in Fig. 2b were then subjected to standard correlation analysis. The auto-correlations of the data show no significant structure, which supports our inference that the individual sequences of land and sea events are randomly distributed. The cross-correlations (Fig. 3a), however, show a strong peak ($>95\%$ confidence) in all three data windows when land quakes lead the sea quakes by 34–38 yr. This is the only peak that is coherent across the entire catalogue; other correlation peaks are not persistent throughout all three data windows.

The degree of correlation is shown in Fig. 2c, where we have re-plotted the land record with a 36-year time-shift. Window B alone is too short to provide a convincing correlation, but window A and its extension to window C clearly reveal a correlation between the land and sea events.

We used a variation of the bootstrap method⁹ of computer-intensive data analysis to see how often such a correlation could be obtained by chance from the available data. Because of the shortness of window B, we focused on window A and shuffled only the land events in this window by assigning new origin times from a random sample of times. Values of the maximum cross-correlation and the corresponding lag for 10,000 random shuffles are shown in Fig. 3b. Nothing anomalous is apparent either at 36 yr or at other lags, indicating that the correlation is

Fig. 3 *a*, Cross-correlations of the filtered land-sea records as a function of land-lead (solid line) and land-lag (dashed line) for windows A, B and C. Light dotted line is drawn at the 95% confidence level in each plot. This level is determined from the formal statistics of the linear correlation coefficient of n degrees of freedom, where n is the number of events in the overlap between the land and sea records being correlated. *b*, Histograms of the maximum cross-correlation and the lag at which it occurred for 10,000 random shuffles of the land (only) events in window A. Arrows mark the values of the observed correlation peak from window A.

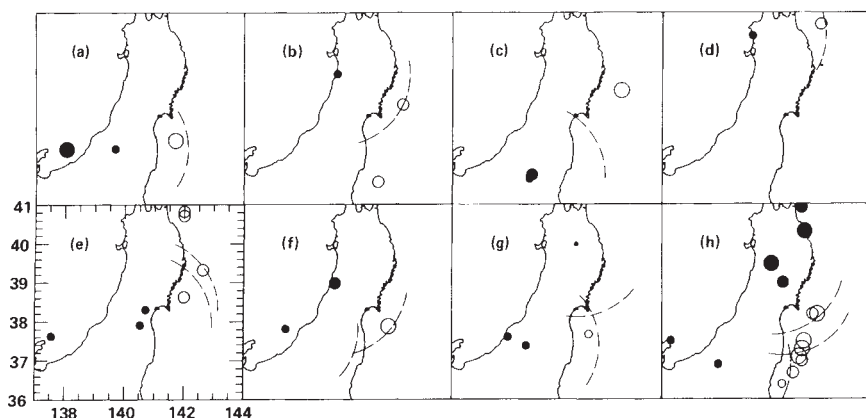
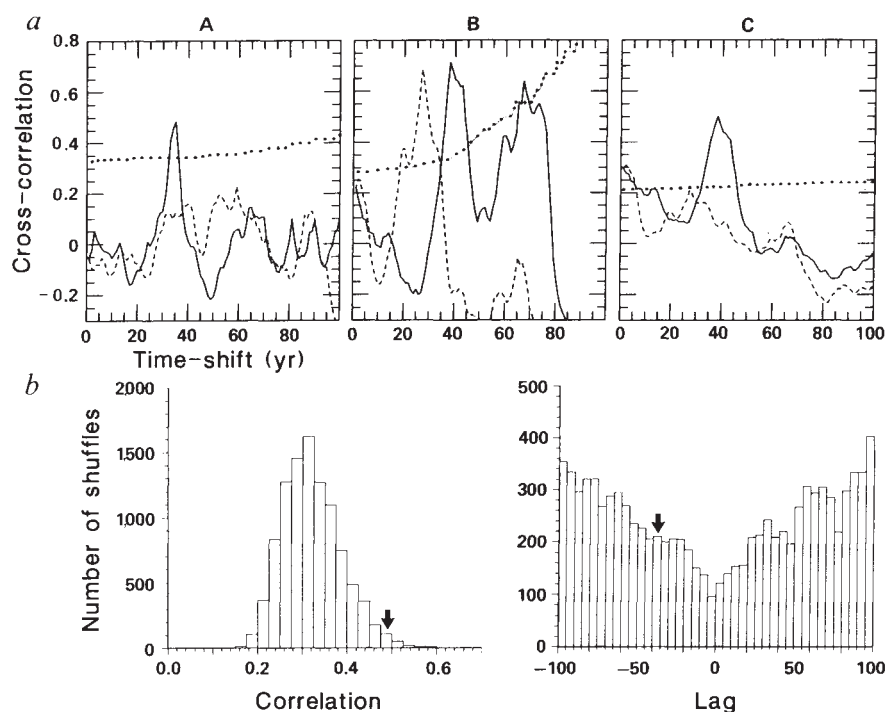


Fig. 4 Location of the land and sea events that produce the correlated pulses (a) to (h) in Fig. 2c. Light dashed arcs are drawn 200 km from the land events, which is the most probable distance in Table 1 between land-sea pairs.

Table 1 Second-order moment analysis of the land-sea records

Distance (km)	Land-lead time* (yr)										Poissonian rate	Probability of largest entry (%)
	58	54	50	46	42	38	34	30	26	22		
50-150	1	-	-	-	-	-	1	-	-	-	0.293	>50
150-250	3	-	-	1	1	2	6	1	1	-	0.384	0.50
250-350	2	-	2	1	1	1	2	3	4	1	0.320	4.1
350-450	-	-	1	-	2	-	2	-	-	1	0.210	21

* The centre time of a 4-yr bin.

not forced by the sea record. The mean of the histogram of the maximum cross-correlation peaks at about 0.30, compared with the observed peak of 0.49, and the spread indicates that only 0.4% of the shuffles exceeds the observed value. Compared with the observed correlation, obtaining a similar result by a chance shuffle is highly unlikely.

Distance between events is included in the analysis by computing the second-order time-space moment^{10,11} of the land-sea records, treating each as an unmarked Poissonian point process. Again, because of the change of earthquake rate noted previously and the relative shortness of window B, we focus our analysis on window A. The second-order moment we calculate,

shown as a matrix in Table 1, is the number of land-sea pairs in a given time-distance interval from the land shock of the pair. Table 1 also lists the Poissonian rates of pair production and the chance probabilities, computed from these rates¹², of obtaining the observed number of land-sea pairs. These rates were determined by second-order moment analysis of random trials (1,000 cases) in which the location boxes in Fig. 1 were filled with random events (in space and time) which had the same first-order Poissonian statistics as the earthquake records. The greatest number of pairs (six) in the Table, with a 0.5% probability of occurring by chance, is found in the time interval 32-36 yr with a mean distance of ~200 km. The spatial correla-

tion (Fig. 4) is not peaked as strongly as the temporal correlation, but this is not surprising because, before seismic instrumentation, the origin times of earthquakes are known more precisely (in terms of the year of occurrence) than the positions of their epicentres.

While subducting, the oceanic plate compresses the continental plate (see inset to Fig. 1). Eventually, the yield strength on land is exceeded, resulting in a large intraplate (land) thrust event which ruptures to relieve the stress and relax the plate. This relaxation, however, is controlled by viscous effects, because the lithosphere is coupled to the asthenosphere¹³. We have solved for the horizontal strain field produced by an edge dislocation in a purely elastic plate over a viscoelastic channel with a Maxwell rheology¹⁴. This model was used originally by Heinze and Sacks¹⁵ to explain post-seismic strain changes measured by geodetic surveys in northern Japan. In our two-dimensional model, the elastic plate is the continental lithosphere, the viscosity in the problem comes from the underlying asthenosphere, and the dislocation is the sudden offset from a thrust earthquake. A thrust source is appropriate because Awata and Kakimi¹⁶ have shown from a study of historical earthquakes that the nature of faulting in this region of Japan is characterized by thrust on large-scale crustal blocks parallel to the island arc. The largest land shock in this region in the modern era (1896 Riku-u) produced ~3 m of thrust displacement.

The viscoelastic model has solutions that are diffuse-like strain pulses, for which the velocity of a pulse (at a given distance from the earthquake) is a function of two unknown parameters: effective elastic thickness H_e of the lithosphere and viscosity η of the asthenosphere. Thatcher *et al.*⁴ used post-seismic uplift data in northern Japan to obtain the value $H_e = 30$ km with less than a 5-km uncertainty, a value in close agreement with that determined by the Research Group for Explosion Seismology¹⁷. Thus in our model, the unconstrained parameter affecting the velocity of pulse propagation is asthenospheric viscosity. The correlation travel-time of 36 yr and mean distance of 200 km between the land-sea events translates into an average pulse velocity of 5.6 km yr^{-1} . Using this velocity, the viscoelastic model gives a viscosity of about $7 \times 10^{18} \text{ Pa s}$, in excellent agreement with a previous estimate⁴ of $10 \times 10^{18} \text{ Pa s}$ in north-east Japan. The velocity is not constant, however; in any diffusive system, the velocity depends strongly on both time and position and these factors must be known to estimate model parameters.

The compression from the motion of the sea plate increases both the normal and shear stress across the locked thrust zone. Normal and shear stress produced in this manner act in opposition to each other: the normal stress controls the static frictional term which locks the fault, whereas the shear stress attempts to drive the opposing plates past each other. Under these conditions, earthquakes in the locked subduction zone may be triggered by external perturbations in the force balance. A large inland earthquake will create a rarefaction in strain which diffuses outward at a rate governed by effective plate thickness and asthenospheric viscosity. As the pulse sweeps past the thrust zone several decades of years later it unlocks the fault by reducing the effective normal stress (or friction) relative to the shear stress, thus triggering an earthquake. This situation is simply the result of viscous retardation of the pulling away of the continental plate from the oceanic plate by the rupture of a decades-old earthquake. We believe the viscoelastic coupling and the simple triggering model has the inherent mechanical stability necessary to explain the persistence and high degree of correlation between the land-sea events in this region of Japan.

5. Melosh, H. J. *J. geophys. Res.* **81**, 5621-5632 (1976).
6. Savage, J. C. & Prescott, W. H. *J. geophys. Res.* **83**, 5005-5007 (1978).
7. Melosh, H. J. *J. geophys. Res.* **83**, 5009-5010 (1978).
8. Yoshii, T. *Tectonophysics* **55**, 349-360 (1979).
9. Diaconis, P. & Efron, B. *Sci. Am.* **248** (5), 116-130 (1983).
10. Kagan, Y. & Knopoff, L. *Phys. Earth planet. Inter.*, **14**, 97-108 (1976).
11. Reasenber, P. *J. geophys. Res.* **90**, 5479-5495 (1985).
12. Beyer, W. H. *CRC Standard Mathematical Tables* 541-544 (CRC Press, Boca Raton, Florida 1984).
13. Elsasser, W. M. in *The Application of Modern Physics to the Earth and Planetary Interiors* (ed. Runcorn, S. K.) 223-246 (Wiley, New York, 1967).
14. Rydelek, P. A. & Sacks, I. S. *Geophys. J.* (submitted).
15. Heinze, W. D. & Sacks, I. S. *Carnegie Inst. Washington Yb.* (1979).
16. Awata, Y. & Kakimi, T. *Earthquake Prediction. Res.* **3**, 231-251 (1985).
17. Research Group for Explosion Seismology *Tectonophysics* **37**, 117-130 (1977).

Anomalous distribution of nuclear and mitochondrial DNA markers in periodical cicadas

Andrew P. Martin* & Chris Simon*†

* Department of Zoology and † Department of General Science, University of Hawaii, Honolulu, Hawaii, USA

Ecological genetics theory identifies the age at first reproduction and life-cycle length as important components of life-history strategies, and predicts that variation in these traits can have significant evolutionary consequences¹. Changes in the timing of reproduction can affect the fitness of an organism², disrupt gene flow³ and lead to speciation⁴. The existence of two life-cycle lengths differing by four years in periodical cicadas of the genus *Magicicada* provides an excellent opportunity to investigate both the control and evolutionary consequences of changes in the timing of reproduction. In *Magicicada*, life-cycle length has been assumed to be a genetically fixed species-specific trait⁵. Alternatively, however, periodical cicadas could switch between the two life-cycle lengths in response to either genetic⁶ or environmental⁷ factors possibly providing a mechanism for the generation of the extant year classes of cicadas⁸, called broods⁹. Here we describe a survey of populations of 13- and 17-year periodical cicadas which suggests that 13-year cicadas in a large region of central North America are descendents of 17-year cicadas that switched their life-cycle length to 13 years, providing the first strong genetic evidence that life-cycle length is a plastic trait. The change in maturation time and the resultant anomalous distribution of cytoplasmic and nuclear gene markers implies that these cicadas became temporally isolated from the parent brood, joined an already existing brood, and thus brought together two independently evolving gene pools.

Magicicada septendecim (17-year) and *M. tredecim* (13-year) are the most widely distributed periodical cicada species. As part of a larger phylogenetic study, 11 populations of 13-year brood XIX cicadas and five populations of 17-year brood X cicadas were surveyed for mitochondrial DNA (mtDNA) restriction endonuclease genotypes, allozyme variation and for a morphological character, abdominal sternite coloration. Allozyme data were collected for three polymorphic enzyme loci— α -glycerolphosphate dehydrogenase, phosphoglucose mutase (PGM) and an esterase—which were known to provide information on brood identity and phylogenetic relationships^{10,11}.

Two distinct mtDNA genotypes were identified, designated A and B. Populations (individuals collected at a given locality) were fixed for one of the two genotypes. Both mtDNA genotypes A and B were present in brood XIX (13-year) cicada populations, whereas all brood X (17-year) populations possessed mtDNA genotype A (Fig. 1a). All individuals with the mtDNA genotype A had similar abdominal sternite coloration and were significantly different from individuals with mtDNA genotype B (*a posteriori* test of homogeneity¹² $P < 0.05$, Fig. 1a). In addition, populations with mtDNA genotype A had different PGM allele frequencies from those with mtDNA genotype B

Received 11 August; accepted 3 October 1988.

1. Cathles, L. M. *The Viscosity of the Earth's Mantle* (Princeton University Press, Princeton, New Jersey, 1975).
2. Peltier, W. R. & Andrews, J. T. *Geophys. J. R. ast. Soc.* **46**, 605-646 (1976).
3. Nur, A. & Mavko, G. *Science* **183**, 204-206 (1974).
4. Thatcher, W., Matsuda, T., Kato, T. & Rundle, J. B. *J. geophys. Res.* **85**, 6429-6435 (1980).

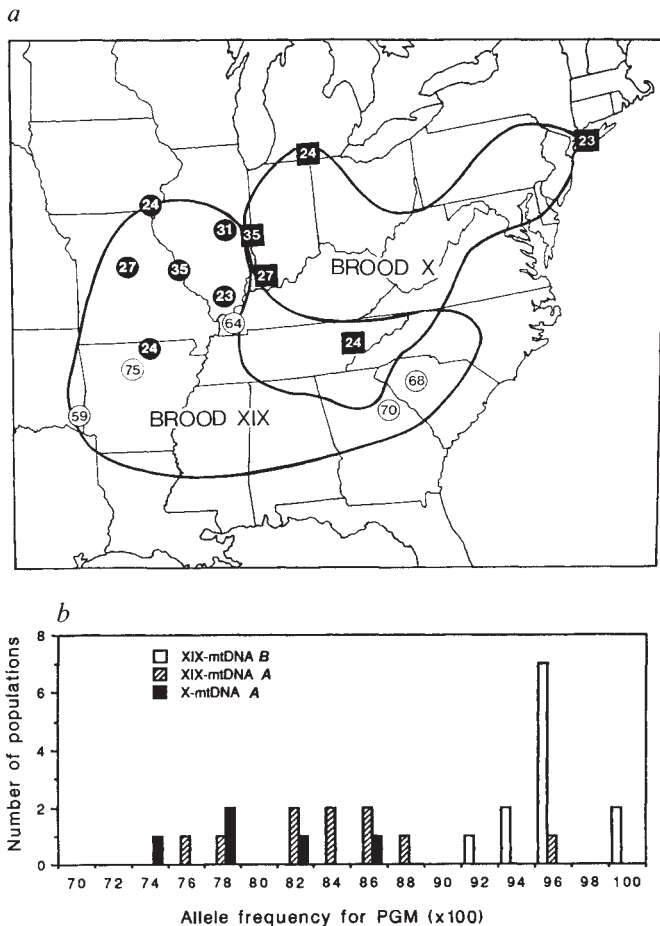


Fig. 1 *a*, Map showing the spatial pattern of mtDNA genotypes and abdominal sternite colouration superimposed on the present-day distribution of broods X and XIX. Brood distributions are enclosed in solid black lines. Circles represent brood XIX (6–12 individuals per population) and squares brood X (4–11 individuals per population) sampled. A total of 118 individuals was sampled. Black circles and squares identify populations with mtDNA genotypes A and white circles represent genotype B. The two distinct genotypes, designated A and B, were identified using fragment lengths for ten restriction enzymes which cut the DNA at two or more sites²⁴. Six enzymes are diagnostic: *Bgl*II, *Bst*NI, *Cla*I, *Eco*RI, *Mbo*I and *Msp*I. The restriction sites for the remaining four enzymes (*Ava*II, *Bcl*I, *Eco*RV and *Xba*I) are identical for all individuals sampled. On the basis of restriction mapping of 50 sites the percentage sequence difference between mtDNA genotype A and B is 2.5 ± 1.0 (ref. 25). The numbers at each locality are a measure of abdominal orangeness. Twenty individuals at each site were assigned a value between one and four (one, mostly black; four, entirely orange), with two and three being intermediate. The population phenotype was calculated by weighting each colour class by the number of individuals in that class and summing the scores. A Mann-Whitney *U*-test established that for abdominal colouration 13-year brood XIX cicadas that have genotype A are not significantly (NS) different from 17-year cicadas with genotype A ($U = 17.5$, NS). These two groups were combined and are significantly different from 13-year cicadas with genotype B ($U = 55$, $P < 0.001$). For more complete distribution records of the broods consult Simon²⁰. *b*, Histogram for the frequency of the most common allele of PGM for 22 populations of brood XIX and five populations of brood X. Forty-eight males and 48 females were sampled at each site. The brood XIX populations are divided into two classes according to mtDNA genotype. As for abdominal colouration, a Mann-Whitney *U*-test shows that populations of 13-year cicadas that have mtDNA genotype A are identical with 17-year cicadas that possess mtDNA genotype A ($U = 38$, NS). These two groups were pooled and are significantly different from 13-year cicada populations with mtDNA genotype B ($U = 172.5$, $P < 0.001$).

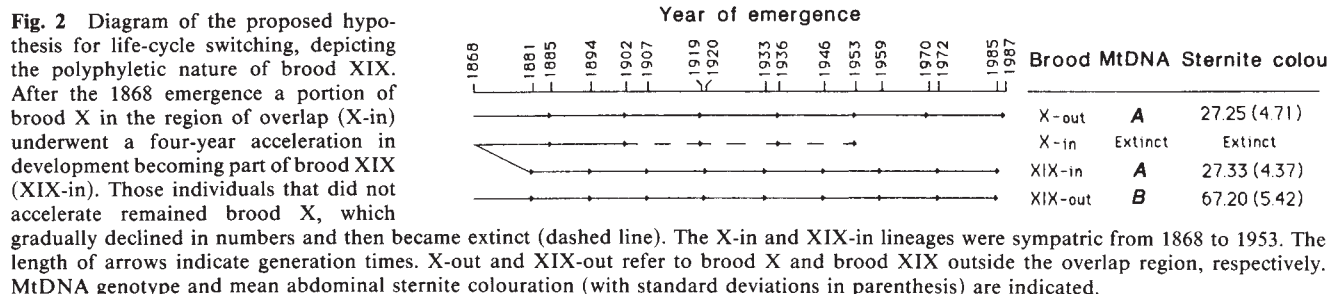
(Fig. 1*b*). These data indicate that for all three characteristics, populations of 13-year cicadas from northwestern Arkansas to southern Iowa and east to the Illinois–Indiana border are more similar to populations of 17-year cicadas than they are to 13-year cicadas collected elsewhere in their range that same year.

There are three reasonable hypotheses which can explain the anomalous mtDNA pattern. The first was proposed in the light of the fact that all 13-year brood XIX cicadas possessing the mtDNA A genotype (characteristic of 17-year brood X cicadas) were collected in a zone of historical overlap of brood X and brood XIX (refs 6, 13, 14). If both broods were present in this region of overlap before 1868 as has been suggested by Lloyd and co-workers⁶, then hybridization between 13- and 17-year cicadas and introgression of genotype A into 13-year cicadas accompanied by elimination of genotype B could be invoked to explain the observed distribution of mtDNA genotypes. Most documented cases of hybridization, however, involve narrow zones in which the distribution of parapatric species come into secondary contact¹⁵. In contrast, for periodical cicadas, if Lloyd and co-workers are correct, there would have been extensive overlap in the distributions enabling hybridization to occur across a wide area. In such a case, given that there have been only eight generations since 1868, the complete elimination of genotype B throughout the entire overlap region would require extremely strong selection favouring individuals with mtDNA genotype A. In addition, if hybridization did occur, we would expect to observe intermediate values for both PGM frequencies and abdominal coloration. The complete lack of intermediacy and the striking abrupt change over a relatively short distance for all three characters suggests that hybridization did not occur.

A second explanation for the observed pattern is that the common ancestor of broods X and XIX was polymorphic for both genotypes A and B and that genotype A was strongly selected in the north whereas genotype B was selected in the south. Again, the lack of intermediacy and the concordant abrupt change observed for all three characters argues against this hypothesis.

A more parsimonious explanation for the genetic identity of 13-year brood XIX cicadas in the overlap zone with 17-year brood X cicadas is the occurrence of a widespread life-cycle switching event in which a large number of 17-year cicadas underwent a four-year acceleration in development to become 13-year cicadas. This life-cycle switching hypothesis can be evaluated by examining the historical emergence records for periodical cicadas in the region of overlap^{6,13,14}. On the basis of these data, we propose that before 1868 northern Arkansas, Missouri, Illinois and southeastern Iowa were exclusively inhabited by 17-year brood X cicadas. After the 1868 emergence, a large portion of the offspring of brood X in these areas underwent a permanent four-year acceleration in development (Fig. 2), appearing in subsequent historical records as brood XIX (13-year) cicadas. The cicadas that did not switch life-cycle length after the 1868 emergence continued to emerge once every 17 years as brood X. The emergence records also indicate that brood X gradually declined to extinction in the region of overlap (Fig. 2; ref. 6).

The decline and local extinction of the brood X cicadas in this region may have been because of competition between 13-year and 17-year cicada nymphs¹⁶ *et al.* Cicada nymphs live in the soil and feed on the xylem fluid of roots¹⁷. Population sizes are enormous—one measurement yielded 1,500,000 individuals per acre¹⁸—indicating that roots may be a limiting resource. Because the second instar growth period for 17-year cicadas is inhibited, 13-year cicadas are larger at the same age and are therefore competitively superior¹⁹. Competition over a number of generations seems to have been sufficient to result in the exclusion of 17-year forms in the region of overlap¹⁶. Indeed, documented cases of 13-year–17-year geographical overlap are rare²⁰.



Other workers have suggested that there are four-year accelerations in development. Lloyd and Dybas⁸ invoked four-year life-cycle shortening as a possible mechanism for the formation of the most widely distributed 17-year broods. Furthermore, observed cases of localized four-year accelerations have been increasingly documented in recent years, whereas life-cycle length changes other than four years are uncommon, and generally involve relatively few individuals²⁰⁻²².

Although our data do not directly address the mechanism that controls life-cycle length switching, they do not favour the hypothesis proposed by Lloyd⁶ *et al.* Their single-locus, two-allele model specifically states that changes in life-cycle length can only occur as a result of hybridization between 13- and 17-year cicadas, whereas our data indicate that life-cycle switching occurs in the absence of hybridization. In addition, all documented cases of cicadas that have been observed to emerge four years ahead of schedule occurred in areas which were inhabited exclusively by 17-year cicadas²⁰.

The actual mechanism that controls life-cycle length switching in periodical cicadas is unknown; however, the trigger for the switch appears to be external. All documented instances of four-year accelerations have occurred in areas of high density, suggesting that life-cycle switching is a density-dependent response to crowding. Perhaps a change in the timing of maturation is analogous to the dormancy strategies shown by annual plants in which some proportion of the available offspring develop at different times to spread the risks in an unpredictable environment²³. Interestingly, the difference in life-cycle length is due to an inhibition or reduction in growth during only the second instar, suggesting that development time can be shortened by skipping this reduced growth period.

The temporal population structure exhibited by year classes of periodical cicadas^{8,20} results in a situation in which a switch in the timing of reproduction can have two simultaneous evolutionary consequences. Such a change in life-cycle length can disrupt gene flow, initiating genetic divergence, and in some cases result in the formation of a new temporally-isolated year class⁸. In addition, life-cycle accelerations can bring together two genetically differentiated gene pools. Because the individuals that switched their life cycle from 17 to 13 years now emerge synchronously with previously isolated brood XIX cicadas, they provide an opportunity for gene flow. In a sense, life-cycle length plasticity is a mechanism for dispersal in time, facilitating gene flow among temporally isolated populations.

We thank James Archie and Steven Palumbi for technical advice. The Cooperative Extensive Service and many other individuals aided in the collection of cicadas. This research was supported by the NSF.

- Lloyd, M. & Dybas, H. S. *Evolution* **20**, 466-505 (1966).
- Marlatt, C. L. *Misc. Res. Work Div. Ent. Bull. U.S.D.A. Div. Ent.* **18**, 52-58 (1898).
- Simon, C. *Syst. Zool.* **28**, 22-39 (1979).
- Archie, J., Simon, C. & Wartenberg, D. *Evolution* **39**, 1261-1274 (1985).
- Sokal, R. R. & Rohlf, F. J. *Biometry*, 776 (Freeman, San Francisco, 1969).
- Haseman, L. U. *Missouri Agric. Exper. Sta. Bull.* **137**, 33 (1915).
- Hyslop, J. A. *U.S.D.A. Bur. Ent. Plant Quar.* **E-364**, 7 (1935).
- Hewitt, G. M. *Trends Ecol. Evol.* **3**, 158-167 (1988).
- Simon, C., Lloyd, M. & Karban, R. *Ecology* **62**, 1525-1535 (1981).
- White, J. & Strehl, C. E. *Ecol. Ent.* **3**, 323-327 (1978).
- Dybas, H. & Davis, D. *Ecology* **43**, 444-459 (1962).
- White, J. & Lloyd, M. *Am. Midl. Nat.* **94**, 127-143 (1975).
- Simon, C. *Proc. Ent. Soc. Am.* **34**, (in the press).
- Kritsky, G. *Proc. Indian Acad. Sci.* (in the press).
- Kritsky, G. *Ohio J. Sci.* **88**, 168-170 (1988).
- Lawlor, L. & Venable, L. *Oecologia* **46**, 272-282 (1980).
- Lansman, R. A., Shade, R. O., Shapira, J. F. & Avise, J. C. *J. Molec. Evol.* **17**, 214-226 (1981).
- Nei, M. & Tajima, F. *Genetics* **105**, 207-217 (1983).

Rate of evolution slowed by a dormant propagule pool

Nelson G. Hairston Jr & Bart T. De Stasio Jr

Section of Ecology and Systematics, Division of Biological Sciences, Cornell University, Ithaca, New York 14853, USA

Theoretical models of plant seed banks and common sense suggest that stores of dormant propagules must slow down the rate of genetic change^{1,2} because they sequester a substantial fraction of the gene pool from the influence of microevolutionary processes in each generation. The magnitude of the effect must represent a balance between the extent of selection and the fraction of the dormant pool that returns to the active population in each generation (and the other microevolutionary processes of migration, drift and mutation). For most populations, in which to observe the action of selection is difficult in itself, detecting the added influence of a dormant propagule pool will be more difficult still. Here we report the effect of hatching of diapausing eggs from lake sediments on the rate of phenotypic change in two populations of the freshwater copepod, *Diaptomus sanguineus*. When the input of new eggs to the sediments was eliminated by a brief but intense two-year appearance of predators, the role of dormant stages in slowing evolution was clearly seen.

D. sanguineus was studied in two small lakes, Bullhead Pond (Bh) and Little Bullhead Pond (LBh) in South Kingstown, Rhode Island. The species is active each year from November to May, making two generations by production of eggs that hatch immediately^{3,4}. In spring, females switch to making diapausing eggs that rest in the lake sediments through the summer. Diapausing eggs begin hatching in November and give rise to the active winter population. The predatory activity of fish (*Lepomis macrochirus*, *L. gibbosus* and *L. auritus*) in Bh and LBh increases in spring as the water temperature rises^{3,5}, and two separate lines of evidence indicate that the copepods' seasonal reproductive pattern permits them to escape the period (spring and summer) of greatest predation. First, the date of the spring switch from production of immediate eggs to diapaus-

Received 17 February; accepted 23 September 1988.

- Stearns, S. C. *Q. Rev. Biol.* **51**, 3-47 (1976).
- Lewontin, R. C. in *The Genetics of Colonizing Species* (eds Baker, H. G. & Stebbins, G. L.) 77-94 (Academic, New York, 1965).
- Dingle, H. & Hegmann, J. P. (eds) *Evolution and Genetics of Life Histories* (Springer, New York, 1982).
- Tauber, C. A. & Tauber, M. J. *A. Rev. Ecol. Syst.* **12**, 281-308 (1981).
- Alexander, R. D. & Moore, T. A. *Misc. Publ. Mus. Zool. Univ. Mich.* **121**, 58 (1962).
- Lloyd, M., Kritsky, G. & Simon, C. *Evolution* **37**, 1162-1180 (1983).
- Lloyd, M. & White, J. *Evolution* **30**, 786-801 (1976).

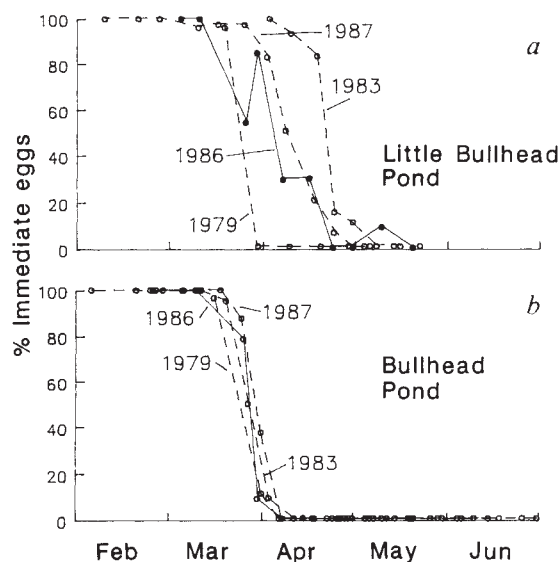


Fig. 1 *a*, Timing of the switch from production of immediate to diapausing eggs by *D. sanguineus* in Little Bullhead Pond before (1979) and after (1983, 1986, 1987) a fish-killing drought in 1981, and before (1983) and after (1986, 1987) two years when predatory fly larvae eliminated maturing copepods. *b*, Timing of the switch for the same years in Bullhead Pond where no fish were killed and fly larvae were never present. Statistical comparisons using logistic regression show significant ($X^2_{1,0.05} > 3.84$) date effects (the inflexion points in Fig. 1 differ) between: BH 1983 versus LBH 1983; BH 1986 versus LBH 1986; BH 1987 versus LBH 1987; LBH 1979 versus LBH 1983, 1986 or 1987; LBH 1983 versus LBH 1986 or 1987; and LBH 1986 versus LBH 1987. Date $\times$ year or date $\times$ pond interactions (slopes of lines in Fig. 1) differ significantly between: BH 1986 versus LBH 1986; BH 1987 versus LBH 1987; LBH 1979 versus LBH 1986, or 1987; LBH 1983 versus LBH 1986 or 1987. Significant date $\times$ pond or date $\times$ year (slope) effects were not observed for: LBH 1979 versus LBH 1983; LBH 1983 versus BH 1979 or 1983. Neither date (inflexion) nor date $\times$ pond or date $\times$ year (slope) effects were significant in the comparisons: BH 1979 versus LBH 1979; BH 1979 versus BH 1983, 1986, or 1987; and the combined data of LBH 1979 and 1983 versus LBH 1986. In Little Bullhead Pond, the response to a relaxation of selection for early diapause to avoid summer fish predation is seen in comparisons of 1979 with 1983, and 1986 with 1987. The effect of the egg bank in retarding response to selection is seen in a comparison of 1983 with 1986 (see text). None of these effects is seen in the Bullhead Pond data. Each point represents an estimate of the fraction of clutches of each egg type carried in the population. Between 15 and 94 clutches of eggs were used to establish each point. Immediate eggs were distinguished from diapausing eggs using previously established methods^{3,4}.

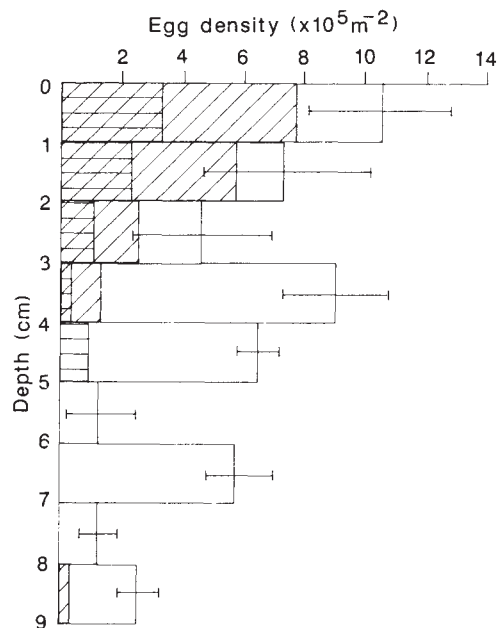


Fig. 2 Depth distribution of diapausing eggs in a sediment core taken from the deepest portion of Little Bullhead Pond on 15 August 1987. The 7-cm-wide core was taken by a diver, divided into 1-cm-thick slices, subsampled in triplicate, and counted at $\times 25$ magnification. Open bars represent all eggs present at each depth interval. Diagonally shaded bars represent eggs judged as viable by the criterion of a visible eye spot. Horizontally shaded bars represent eggs shown to be viable by hatching in the laboratory^{3,4}. Error bars are ± 1 s.e. in triplicate subsamples at each depth in the core.

ing eggs agrees very closely with the theoretically optimal switch date predicted from a mathematical model based on field measurements of the onset of predation^{3,6}. Second, a natural fish removal event (1981 drought) in LBh resulted in a shift in the timing of the date of switch to diapause to significantly later in the spring (Fig. 1a). Such a shift would be expected from a relaxation of the selection pressure to enter diapause to avoid fish predation⁷⁻⁹. In Bh, where no fish kill occurred, the timing of diapause remained unchanged over the same period (Fig. 1b). Other explanations for the shift in diapause timing (immigration and physiological response) are not supported by data: copepod abundance after the drought greatly exceeded plausible input from dispersal, and copepods did not change diapause timing when reared in a fishless pond^{4,7}. In addition, a rapid response to selection is consistent with high laboratory-measured heritability of diapause response to photoperiod in the LBh population: $h^2 = 0.48$ (0.14 s.e.). Heritability was estimated with a threshold selection design¹⁰, and replicated four-

times using a total of 1,051 mothers and daughters (N.G.H. and T. A. Dillon, manuscript in preparation).

The diapausing eggs produced by *D. sanguineus* do not all hatch in the season following their production. Some are stored in the lake sediments for longer periods with the result that a large pool of dormant propagules accumulates. Although dormant copepod eggs have been studied in aquatic sediments^{11,12}, the frequency of egg survival beyond one year has not. Thus pools of dormant propagules, though potentially common among the Copepoda, have not previously been described. Sediment cores taken in both lakes (7 in Bh and 1 in LBh) reveal viable egg densities ranging from 6×10^4 to 5×10^5 eggs m^{-2} near the surface, with a general reduction in abundance at greater depths (Fig. 2). Viable eggs were found in all cores down to 5 cm below the sediment surface, and occasionally deeper. Sediment accumulation rates are not known for these lakes, but sedimentation rates of 0.05 to 0.35 cm yr^{-1} in a broad variety of other northern temperate lakes¹³⁻¹⁷ suggest that viable eggs at 5 cm should be between 15 and 100 yr old. The actual age of buried eggs must also depend on rates of sediment reworking by benthic organisms^{11,18} and physical processes^{15,19}, which for these lakes are unknown. Nevertheless, two lines of evidence show that buried eggs remain viable and continue to hatch for at least three years after they are produced. First, 24 inverted funnel traps were placed over the sediment in Bullhead Pond with the open ends buried 1–2 cm and left in place for up to three consecutive years. A plastic flask containing filtered pond water was placed over the stem of each funnel and changed weekly during the ice-free season. The traps yielded 9.1×10^3 emerging nauplii m^{-2} ($\pm 5.7 \times 10^3$ s.e., $n = 3$) the first year and 1.1×10^3 ($\pm 5.8 \times 10^2$, $n = 3$), and 6.4×10^3 ($\pm 3.2 \times 10^3$, $n = 2$) nauplii m^{-2} in subsequent years. Second, after the fish-killing drought in LBh, predatory fly larvae (*Chaoborus americanus*) became extraordinarily abundant (up to 41 l^{-1}), whereas they

Table 1 Maximum numbers of copepods (first stage nauplii, adults, and newly produced eggs) and predatory fly larvae (*Chaoborus*) per litre (1 s.d.) in Little Bullhead Pond during winter and spring cohorts

Year	Season	Nauplius I (Numbers litre ⁻¹)	Adults (Numbers litre ⁻¹)	New eggs (Numbers litre ⁻¹)	<i>Chaoborus</i> (Numbers litre ⁻¹)	Fish
1977-1978	winter	no data	no data	no data	no data	present
1978	spring	2.53 (2.30)	7.15 (1.41)	17.1 (8.8)	0.0	present
1978-1979	winter	1.26 (0.48)	0.92 (0.57)	0.22 (0.02)	0.0	present
1979	spring	0.43 (0.01)	0.24 (0.11)	0.10 (0.04)	0.0	present
1979-1980	winter	no data	no data	no data	no data	present
1980	spring	3.09 (1.29)	2.75 (2.62)	17.2 (1.6)	0.0	present
1980-1981	winter	0.96 (0.43)		pond dry		absent
1981	spring			pond dry		absent
1981-1982	winter			pond dry		absent
1982	spring	1.95 (0.21)	4.05 (2.25)	33.1 (11.2)	0.02 (0.03)	absent
1982-1983	winter	0.20 (0.13)	0.02 (0.01)	0.11 (0.15)	0.39 (0.55)	absent
1983	spring	0.71 (0.21)	0.40 (0.37)	0.40 (0.37)	1.16 (0.39)	absent
1983-1984	winter	0.15 (0.07)	0.02 (0.005)	0.01 (0.01)	0.48 (0.12)	absent
1984	spring	0.03 (0.01)	0.01 (0.01)	0.00 (—)	4.15 (1.07)	absent
1984-1985	winter	0.01 (0.01)	0.00 (—)	0.00 (—)	1.40 (1.13)	absent
1985	spring	0.86 (0.39)	0.002 (0.003)	0.00 (—)	3.79 (0.17)	absent
1985-1986	winter	0.03 (0.04)	0.07 (0.09)	0.43 (0.18)	0.01 (0.02)	absent
1986	spring	1.09 (0.76)	0.14 (0.20)	5.97 (4.77)	0.005 (0.01)	absent
1986-1987	winter	0.14 (0.04)	0.75 (1.06)	3.47 (3.37)	0.07 (0.21)	absent
1987	spring	1.34 (1.61)	4.12 (3.47)	32.0 (24.8)	0.001 (0.001)	absent

Duplicate zooplankton samples were collected weekly using a Clarke-Bumpus quantitative plankton sampler fitted with a 75- μ m-mesh net. Samples were preserved in buffered formalin and subsamples were counted at $\times 25$ magnification. The periods when planktivorous sunfish (*Lepomis auritus*) were present and absent are also shown. When present, these fish were extremely dense (2.5 fish m⁻³). Both ponds contained planktivorous sunfish for at least the 16 yr preceding the 1981 drought. *Chaoborus*, which appeared after the fish were killed by drought, severely depressed copepod abundance in 1984 and 1985. The copepod population recovered after the fly larvae again became scarce in 1986 and 1987.

had previously been absent (Table 1). *C. americanus* is a large zooplankton that is excluded from lakes with planktivorous fish²⁰. Capable of consuming between 2 and 10 copepods per larva per day²¹⁻²³, these predators decimated the emerging population of *D. sanguineus* for two consecutive years (1983-1984 and 1984-1985). Each autumn nauplii appeared in the water column from the hatching of diapausing eggs, but were eliminated before reproducing (Table 1). In summer 1985 the *Chaoborus* population fell drastically (presumably as a result of starvation), and in the following winter (1985-1986) *D. sanguineus* reappeared from the sediments, matured, and successfully reproduced. Thus, drawing only on diapausing eggs produced in 1982-1983 and before, copepods appeared in the active population in 1983-1984, 1984-1985, and 1985-1986, indicating that there was at least a three-year egg pool.

The influence of this egg pool on the evolutionary dynamics of the active population is reflected in the diapause timing of the copepods reproducing in 1986 (Fig. 1a). Instead of mirroring the timing present in 1983, *D. sanguineus* switched from producing immediate eggs to diapausing eggs at a time distinctly intermediate between 1979, when fish were present, and 1983, after fish removal. We interpret these events as follows: (1) most of the eggs produced in 1983 hatched in 1983-1984 and 1984-1985 and were eliminated by *Chaoborus*, but a significant fraction did not hatch until 1985-1986 (in 1982, LBH refilled late in the autumn, and the copepods did not mature until well past the usual date of the spring switch to diapause⁷), (2) some eggs produced before the fish-kill (1981) also hatched from the dormant pool in 1985-1986, and (3) although each year's contribution of eggs hatching from the sediments was presumably low, the net effect of many years of diapausing eggs from pre-fish-kill, combined with only two post-fish-kill years, resulted in a roughly equal contribution to the active copepod population in 1985-1986. The substantially broader distribution of switching times (giving significantly different slopes in Fig. 1) seen in 1986 in comparison with 1979 and 1983, supports the interpretation that hatching in 1985-1986 came from a mixture of phenotype distributions.

After fish were killed by the drought in Little Bullhead Pond, selection favoured those individuals that continued to make immediate eggs late into the spring, as the change in the timing of diapause from 1979 to 1983 shows^{7,8}. The effect of eggs hatching from the dormant pool is to slow down the rate of change. It is made visible by the insertion of two years when the copepods failed to recharge the sediments successfully with new diapausing eggs. The *Chaoborus* population remained very low in 1987, and the timing of the switch to diapause again advanced to later in the spring, as expected from the continued action of natural selection (Fig. 1a).

How can selection act rapidly to shift the timing of diapause, while at the same time being slowed by eggs hatching from the dormant pool? The result depends on the dynamics of the eggs stored in the sediments. If there is much more hatching of the diapausing eggs during the year immediately following their production than in subsequent years, then the dormant pool will generally only slightly retard selection. Eggs that fail to hatch in the year after they are produced, even though potentially quite abundant, will be increasingly buried by sediment, and so will be less and less likely to receive the hatching stimulus in later years. This process is particularly probable for dormant eggs in lake sediments, in which the frequency of stratum disturbance is patchily distributed and declines markedly with depth^{11,18}.

For *D. sanguineus* in Little Bullhead Pond, the dormant propagule pool had the extremely important role of keeping the population present in the pond through two years (1983-1984 and 1984-1985) of completely failed reproduction. Natural selection for this ability may ultimately underlie the existence of the egg pool, or the pool may simply result from the physical process of egg burial by sedimentation. In either case, the effect of emergence from the dormant pool on the microevolutionary dynamics of the active population was profound.

We thank E. J. Olds and T. A. Dillon for assistance; the Department of Zoology, University of Rhode Island for space and logistic support; J. E. O'Brien and R. P. Clark for access to the lakes, and C. E. McCulloch for statistical advice. Useful

comments were made on the manuscript by A. R. Templeton, J. B. McGraw, W. A. Searcy, N. G. Hairston Sr, K. D. Hambright and S. F. Tjossem. The research was supported by NSF grants to N.G.H. and a Sigma Xi grant to B.T.D.

Received 15 August; accepted 3 October 1988.

1. Templeton, A. R. & Levin, D. A. *Am. Nat.* **114**, 232–249 (1979).
2. Brown, J. S. & Venable, D. L. *Am. Nat.* **127**, 31–47 (1986).
3. Hairston, N. G. Jr & Munns, W. R. Jr *Am. Nat.* **123**, 733–751 (1984).
4. Hairston, N. G. Jr & Olds, E. J. *Oecologia* **61**, 42–48 (1984).
5. Hairston, N. G. Jr, Walton, W. E. & Li, K. T. *Limnol. Oceanogr.* **28**, 935–947 (1983).
6. Hairston, N. G. Jr, Olds, E. J. & Munns, W. R. Jr *Verh. int. Verein. theor. angew. Limnol.* **22**, 3170–3177 (1985).
7. Hairston, N. G. Jr & Walton, W. E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4831–4833 (1986).
8. Hairston, N. G. Jr in *Predation: Direct and Indirect Impacts on Aquatic Communities* (eds Kerfoot, W. C. & Sih, A.) 281–290 (University Press New England, Hanover, New Hampshire, 1987).
9. Taylor, F. & Spalding, J. *Am. Nat.* **131**, 678–699 (1988).
10. Falconer, D. S. *Introduction to Quantitative Genetics* 2nd edn (Longman, New York, 1981).
11. Marcus, N. H. *Mar. Evol. Prog. Ser.* **15**, 47–54 (1984).
12. Champeau, A. *Ann. Fac. Sci. Marseille* **44**, 155–189 (1970).
13. Davis, M. B. *Ecology* **50**, 409–421 (1969).
14. Griffiths, M. & Edmondson, W. T. *Limnol. Oceanogr.* **20**, 945–952 (1975).
15. Lehman, J. T. *Quat. Res.* **5**, 541–550 (1975).
16. Davis, R. B. & Anderson, D. S. *Hydrobiologia* **120**, 69–87 (1985).
17. Moritz, C. *Hydrobiologia* **145**, 309–314 (1987).
18. Davis, R. B. *Limnol. Oceanogr.* **19**, 342–346 (1974).
19. Likens, G. E. & Davis, M. B. *Verh. int. Verein. theor. angew. Limnol.* **19**, 982–993 (1975).
20. von Ende, C. N. *Ecology* **60**, 119–128 (1979).
21. Fedorenko, A. Y. *Limnol. Oceanogr.* **20**, 250–258 (1975).
22. Pastorok, R. A. in *Evolution and Ecology of Zooplankton Communities* (ed. Kerfoot, W. C.) 538–554 (University Press New England, Hanover, New Hampshire, 1980).
23. Walton, W. E. *J. Plank. Res.* **10**, 110–114 (1988).

A copia-like transposable element family in *Arabidopsis thaliana*

Daniel F. Voytas & Frederick M. Ausubel

Department of Genetics, Harvard Medical School and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

The fast generation time, small genome size and extensive genetic map of the crucifer *Arabidopsis thaliana* have made it the subject of an increasing number of studies in plant molecular genetics¹. As transposable elements have greatly facilitated genetic analysis in a variety of species, we have attempted to identify an endogenous *A. thaliana* transposable element. We report here the discovery of a family of such elements, which we refer to as Ta1 elements. Sequence analysis of one such element shows that it is closely related to retrotransposons and integrated retroviral proviruses, being bound by a direct sequence repeat and having an open reading frame with clear sequence similarity to the polyprotein of the *Drosophila melanogaster* retrotransposon *copia*. The sequence of an empty target site of a Ta1 element shows that insertion is accompanied by a five-base-pair target-site duplication and that Ta1 has transposed in the period of time since divergence of two races of *A. thaliana*.

A number of transposable elements have been initially identified as restriction fragment length polymorphisms (RFLPs) in comparisons of cloned DNA sequences between strains or natural isolates of a species (for example Ty1 in yeast² and Tc1 in nematodes³). We undertook a screen of RFLPs among 16 geographical races (ecotypes) of *A. thaliana* to identify insertional polymorphisms that could have arisen by DNA transposition.

During our RFLP analysis, we discovered a 2.3 kilobase (kb) insertional polymorphism in the ecotype Kas-1 (data not shown), located approximately 5 kb downstream of the chalcone synthase gene, which had been previously cloned in our laboratory from the Landsberg ecotype⁴. By cloning the Kas-1 insertion and using this sequence as a hybridization probe in Southern

blotting experiments, we found that the insertion was a member of a low-copy (one to three copies in most of the ecotypes analysed), interspersed family of sequences which we called Ta1 (data not shown). The Kas-1 insertion (Ta1-1) was found to contain an internal 2.9 kb deletion compared to other Ta1 copies (data not shown). We therefore cloned two additional Ta1 elements from the Landsberg ecotype (La-0), both of which were 5.2 kb and structurally similar on the basis of restriction endonuclease maps of these sequences.

We determined the complete nucleotide sequence of one of the two 5.2 kb Landsberg Ta1 elements (Ta1-3). The overall structure of Ta1-3 is similar to retrotransposons and integrated retroviral proviruses⁵ (Fig. 1). Ta1-3 has a large internal domain of 4,190 base pairs (bp) bounded by 514-bp long terminal direct repeats (LTRs). The LTRs contain short 4-bp inverted repeats at their ends, and the two-nucleotide LTR end sequences (5' TG...CA 3') are the same as those found at the ends of retroviral and retrotransposon LTRs. A 5-bp direct repeat flanks the Ta1-3 insertion. To determine whether this repeat was the consequence of transposition, we cloned and sequenced the corresponding empty target site of Ta1-3 from the ecotype Columbia (Col-0) which does not carry a copy of this element. As shown in Fig. 1, the 5-bp repeats flanking Ta1-3 represent a duplicated target sequence, suggesting that Ta1-3 transposes by a mechanism similar to other retrotransposons and has done so since the two races separated.

First- and second-strand DNA synthesis of retroviral-like elements is primed from sites located within the internal element domain immediately adjacent to the LTRs⁵. These priming sites consist of a short region of transfer RNA homology at the 5' end of the internal domain and a purine-rich sequence at the 3' end. Starting two base pairs within the internal domain adjacent to the 5' LTR of Ta1-3 is a 14 bp sequence identical to the 3' ends of two plant initiator methionine tRNAs (Fig. 1). Similarly, adjacent to the 3' LTR of Ta1-3 is a purine-rich sequence that may serve to prime second-strand DNA synthesis.

The internal 4.2 kb region of Ta1-3 contains a single large open reading frame (ORF) that encodes a presumptive 1,291 amino-acid polyprotein. This ORF is similar to the protein products of retrotransposons and retroviruses (Figs 2 and 3) across conserved amino-acid domains that include: an RNA-binding domain presumably involved in encapsidating element RNA into viral or viral-like particles^{6,7}; a protease domain that cleaves the polyprotein into separate functional polypeptides⁸; an integrase domain that is required for the integration of the respective elements into the host genome^{9,10}; and a reverse transcriptase domain which converts RNA transcripts from the element into DNA¹¹. It has previously been noted that the order of the integrase and reverse transcriptase domains in *copia* and the yeast retrotransposon, Ty1, (*int-rev*) is different from that in other characterized retrotransposons and retroviruses (*rev-int*)¹². The *int-rev* domains appear in the same order in Ta1-3 as in *copia* and Ty1 (Fig. 2).

In addition to the order of the protein domains, Ta1-3 and *copia* share many other features. First, Ta1-3 and *copia* both cause 5-bp target site duplications¹³ and share homology with tRNA^{met} at the 5' end of their internal domains¹⁴. Second, Ta1-3 and *copia* are the only characterized retroviral-like elements that encode a single polyprotein; most retrotransposons and retroviruses encode two or more overlapping reading frames separated by a frame shift⁵ (for example Ty1; see Fig. 2). Third, the amino-acid similarity between Ta1-3 and *copia* extends beyond the conserved amino-acid residues that characterize each of the protein domains (Fig. 3). This is particularly true for the integrase and reverse transcriptase; the 250 amino acids that encompass each of these regions share 37% amino-acid identity (Fig. 2). The extent of amino-acid sequence identity between Ta1-3 and *copia* and the similarities in their structural organization suggest that these two elements comprise a separate, closely related class of retrotransposons.

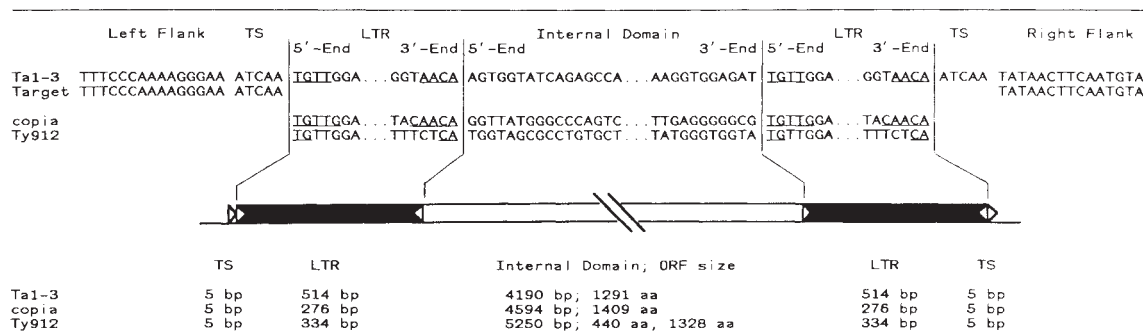


Fig. 1 Structural features of Ta1-3, copia and Ty912. Ta1-3 flanking sequences are aligned with the empty target site from the *A. thaliana* race Columbia. Inverted repeat sequences at the ends of the long terminal repeats (LTRs) are underlined. Putative primer sites at the ends of the internal domain are shown; 5'-end sequences of Ta1-3 shown are identical to tRNA_i^{met} of bean and wheat¹⁵. Structural features of the *D. melanogaster* copia element¹² and the yeast element, Ty912¹⁶ have been previously described. TS, target site duplication; aa, amino acids. The diagram depicts the structural features of retrotransposons; arrowheads outside the element signify target site duplications; black boxes depict the LTRs; arrowheads inside the boxes represent the LTR inverted repeat end sequences.

Methods. *A. thaliana* ecotypes were obtained from Dr A. R. Kranz of the *Arabidopsis* Information Services with the exception of Mv-0 which was isolated from a wild *A. thaliana* population growing on Martha's Vineyard, Massachusetts, USA. The ecotypes represent wild populations from the following locations¹⁷: La-0, West Germany; Col-0, West Germany; Kas-1, India; Co-4, Portugal; Sei-0, Italy; No-0, East Germany; Mv-0, United States; Li-0 Spain; Cvi-0, Cape Verde Islands; Fi-3, Finland; Ba-1, Great Britain; Hau-0, Denmark; Aa-0, West Germany; Ge-0, Switzerland; Ms-0, Soviet Union; Ag-0, France; Mh-0, Poland. To identify RFLPs, partially characterized genomic clones of single-copy *A. thaliana* DNA (pOCA:PAL5 (R. Feinbaum, unpublished), λ CHS2; ref. 4; λ At3102; ref. 18) were used as hybridization probes for Southern filters containing ecotype DNAs. Together, these clones hybridized to genomic fragments encompassing 100 kb of the *A. thaliana* genome (~0.14%; ref. 19). Single and double digests of ecotype DNAs with *Bam*HI and *Bgl*II were used to distinguish restriction site gain or loss from insertional polymorphisms. Plant DNA preparation and analysis methods were similar to those described²⁰. The Ta1-1 insertion was cloned from a library constructed in pUC12 with size-selected Kas-1 DNAs digested to completion with *Xba*I. Desired recombinants were identified by colony hybridizations²⁰ using a DNA probe from λ CHS2. Additional Ta1 element copies were cloned from a complete La-0 genomic library constructed in lambda FIX (Stratagene) and propagated on *Escherichia coli*, ER1458 (ref. 21). Recombinant phage were identified using a DNA probe from Ta1-1. The Col-0 target site was cloned in the vector lambda ZAPII (Stratagene) using *Hind*III-digested, size-selected Col-0 DNA. Cloned DNA which flanks the Ta1-3 insertion was used as a hybridization probe to identify desired recombinants. The DNA sequence of Ta1-3 was obtained for both strands by the dideoxy method using modified T7 polymerase²⁰ (U.S. Biochemical Corp.). Oligonucleotides used to prime sequencing reactions were synthesized on a Bioscience DNA synthesizer (New Brunswick Scientific). These sequence data will appear in the EMBL/GenBank/DBJ Nucleotide Sequence Databases under the accession number X13291.

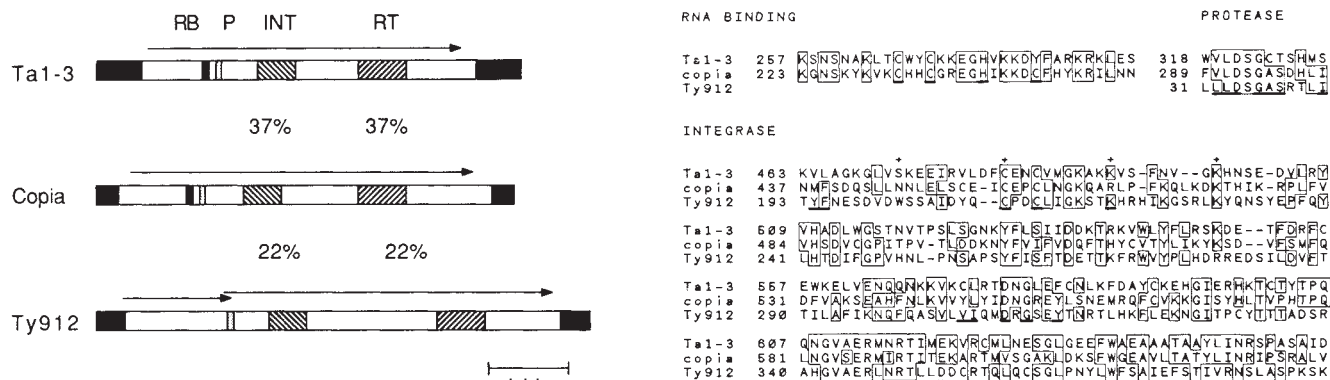
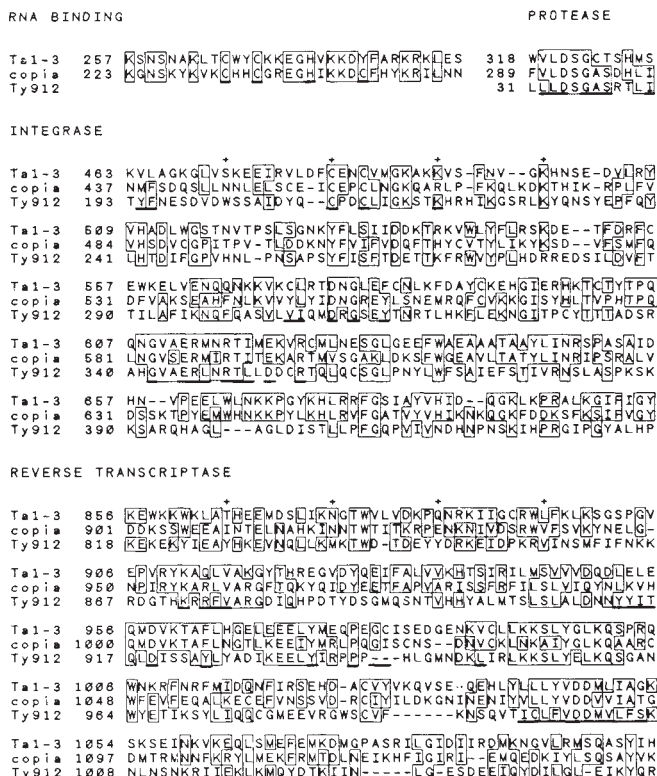


Fig. 2 Organization of open reading frames and conserved amino-acid sequence domains among Ta1-3, copia and Ty912. Elements are drawn to scale with the LTRs represented by black boxes at the ends of the elements. Arrows above the elements depict open reading frames. Boxes within the internal portion of the element represent conserved amino-acid domains: RB, RNA-binding domain; P, protease domain; INT, integrase domain; RT, reverse transcriptase domain. Percentages reflect the amino-acid sequence identity between the integrase and reverse transcriptase domains (Fig. 3). Values for comparisons between Ta1-3 and Ty912 are: integrase, 23%; reverse transcriptase, 24%. Ty912 does not encode an RNA-binding domain⁶.

The original Ta1 insertion was identified by RFLP analysis of only 100 kb of *A. thaliana* DNA (see Fig. 1 legend). This suggests that transposable elements may be found throughout the *A. thaliana* genome, and that transposable element activity may be a major source of genetic variability in this species.

Ta1-3 is the first plant retroviral-like element to show a high degree of amino-acid similarity to conserved domains found within the protein products of retrotransposons and retroviruses. The extent of conservation between Ta1 and copia further



demonstrates that retrotransposons are common to both plants and animals and that these elements may share a common ancestral origin.

We would like to thank F. Winston, S. Rodermel and members of the Ausubel lab for helpful discussions throughout this study. This work was supported by a grant from Hoechst AG to Massachusetts General Hospital.

Received 30 August; accepted 6 October 1988.

1. Meyerowitz, E. M. *A. Rev. Genet.* **21**, 93-111 (1987).
2. Cameron, J. R., Loh, E. Y. & Davis, R. W. *Cell* **16**, 739-751 (1979).
3. Emmons, S. W., Yesner, L., Ruan, K. & Katzenberg, D. *Cell* **32**, 55-65 (1983).
4. Feinbaum, R. & Ausubel, F. M. *Molec. cell. Biol.* **8**, 1985-1992 (1988).
5. Varmus, H. in *Mobile Genetic Elements* (ed. Shapiro, J. A.) (Academic, New York, 1983).
6. Covey, S. *Nucleic Acids Res.* **14**, 623-633 (1986).
7. Fuetterer, J. & Hohn, T. *Trends biochem. Sci.* **12**, 92-96 (1987).
8. Toh, H., Ono, M., Saigo, K. & Miyata, T. *Nature* **315**, 691 (1985).
9. Chiu, I.-M., Callahan, R., Tronick, S. R., Schlom, J. & Aaronson, S. A. *Science* **223**, 364-370 (1984).
10. Panganiban, A. T. & Temin, H. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7885-7889 (1984).
11. Patarca, R. & Haseltine, W. A. *Nature* **309**, 288 (1984).
12. Mount, S. M. & Rubin, G. M. *Molec. cell Biol.* **5**, 1630-1638 (1985).
13. Dunsin, P., Brorin, W. J., Simon, M. A. & Rubin, G. M. *Cell* **21**, 575-579 (1980).
14. Kikuchi, Y., Ando, Y. & Shiba, T. *Nature* **323**, 824-826 (1986).
15. Gauss, D. H. & Sprinzl, M. *Nucleic Acids Res.* **11**, r1-r53 (1983).
16. Clare, J. & Farabaugh, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2829-2833 (1985).
17. Kranz, A. R. *Arabidopsis Inf. Serv.* **18**, 173-176 (1981).
18. Chang, C. & Meyerowitz, E. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1408-1412 (1986).
19. Leutwiler, L. S., Hough-Evans, B. R. & Meyerowitz, E. M. *Molec. Gen. Genet.* **194**, 15-23 (1986).
20. Ausubel, F. M. *et al. Current Protocols in Molecular Biology* (Greene Publishing Associates/Wiley-Interscience, New York, 1987).
21. Raleigh, E. A. & Wilson, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9070-9074 (1986).
22. Needleman, S. B. & Wunch, C. D. *J. molec. Biol.* **48**, 443-453 (1970).
23. Stucka, R., Hauber, J. & Feldmann, H. *Current Genetics* **11**, 193-200 (1986).

Expression of human proteoglycan in Chinese hamster ovary cells inhibits cell proliferation

Yu Yamaguchi & Erkki Ruoslahti

Cancer Research Center, La Jolla Cancer Research Foundation,
10901 North Torrey Pines Road, La Jolla,
California 92037, USA

In studying the functional role of an extracellular matrix proteoglycan, decorin, we have made observations that suggest a role for this proteoglycan in the control of cell proliferation. Extracellular matrices are made up of different combinations of collagens, elastin, hyaluronic acid, proteoglycans and various glycoproteins such as fibronectin. Most of these components can interact with cells, and much of the control of cell adhesion, migration and differentiation appears to be mediated by these interactions¹. Earlier studies have also attributed growth-regulatory activities to intact extracellular matrices², but the individual molecules responsible for these effects have not been characterized. We report here that Chinese hamster ovary (CHO) cell lines expressing human decorin³⁻⁹ from a stably transfected complementary DNA construct form a more orderly monolayer and grow to a lower saturation density than control cells lacking decorin. The extent of the morphological changes correlates with the level of decorin expression, and the saturation density is inversely proportional to it. The reduction in the saturation densities of the cell lines with the highest expression of decorin is more than 50%. These results reveal a novel growth inhibitory mechanism which may be related to contact inhibition of cell proliferation.

Decorin is a fibroblast proteoglycan with a core protein of relative molecular mass (M_r) 40,000 to which is attached a single chondroitin sulphate/dermatan sulphate chain³⁻⁶. The core-protein amino-acid sequence has been determined by cDNA cloning^{7,8}. Decorin was expressed in dihydrofolate reductase (dhfr)-deficient CHO cells by cotransfection with the expression

vectors pSV2-dec, in which a full-length decorin cDNA⁷ was cloned into pSV2¹⁰, and pSV2dhfr¹¹ as a selection marker. This system permits coamplification of the sequences coding for decorin and dhfr upon exposure to increasing concentrations of methotrexate (MTX)¹². The primary transfectants expressed decorin in amounts barely detectable by ³⁵S₄-labelling and immunoprecipitation with an antipeptide antibody specific for decorin. Amplification of the integrated pSV2-dec sequences by MTX selection of the primary transfectants resulted in the selection of cell lines with increasing levels of decorin expression (Fig. 1a and b). The highest expressor cell lines produced up to 25 pg decorin per cell in the 24-hour period (Table 1). No decorin could be detected in the original CHO cells, or in control cell lines prepared by transfecting with the vector alone followed by gene amplification. Electrophoretic analysis of culture media labelled with [³H]leucine (Fig. 1c) showed that fibronectin secretion was not substantially altered in the transfected cells.

Examination of the appearance of the transfected cells revealed a change in the morphology of decorin-expressing cells. Each of a total of eight independent cell lines expressing substantial amounts of decorin after amplification consistently appeared to spread more on the substrate and formed a more even monolayer than the unamplified or amplified control cells (Fig. 2). Quantitation of the average surface area of the cells in low-density cultures showed that the decorin-expressing cells occupied up to nearly three times the surface area of the control lines. In addition, the increase in area was proportional to the amount of decorin synthesized (Table 1). The low-level expressors had a morphology that was indistinguishable from the control cells and showed no significant increase in surface area.

The morphological appearance of the decorin-expressing cells suggested that these cells became quiescent at a low saturation density. Indeed, quantitation of their growth parameters showed that they grew more slowly and became arrested in their growth at lower saturation densities than control cell lines (Fig. 3a). A plot of the saturation density versus decorin production for all the cell lines isolated in this study showed that the increased expression of decorin inversely correlated with the saturation density of these cell lines (Fig. 3b).

Proteoglycans secreted by cells have previously been found to inhibit cell attachment¹³⁻¹⁵, and the proteoglycan preparations with such an activity include a mixture of two proteoglycans, one of which is decorin¹⁵. The changes in the decorin-expressing cells we document here are therefore unexpected. Differences in the assay conditions may account for this apparent discrepancy, because our decorin-producing cell lines continuously synthesize their own decorin during the processes of cell attachment and growth, whereas previous studies examined the effects of soluble decorin on the attachment of cells to a fibronectin substrate.

Most of the decorin produced by cells in culture is secreted in the culture medium, but a small proportion of decorin appears to be associated with the cell surface⁹. This latter fraction may be important to the effects of decorin on cell attachment for two reasons. This fraction may be about to be internalized by the cell, and this could be important because heparin and heparan sulphate taken up by cells have growth-inhibitory activity^{16,17}. Alternatively, the expression of decorin could affect the availability of growth factors to cells because many growth factors are known to bind to proteoglycans through the glycosaminoglycan component, heparan sulphate. However, it is not known whether chondroitin sulphate and dermatan sulphate, the glycosaminoglycans or decorin, share these activities.

Decorin can also bind to extracellular matrix components such as type-I collagen and fibronectin^{9,18,19} and this provides a possible alternative mechanism for the decorin effects on cell morphology and proliferation. Decorin could provide an additional link between cell surface and extracellular matrix, as

Fig. 1 Expression of recombinant decorin in CHO cells. Proteins were separated by SDS-polyacrylamide electrophoresis in 7% gel under reducing conditions and detected by fluorography. *a*, $^{35}\text{SO}_4$ -labelled culture supernatants immunoprecipitated with rabbit anti-peptide antiserum prepared against the NH_2 -terminus of human decorin⁷. *b*, Total $^{35}\text{SO}_4$ -labelled products secreted into culture medium. *c*, Total [^3H] leucine labelled products secreted into culture medium. FN, fibronectin. Lane 1, control transfected cell line A, an unamplified cloned cell line transfected with pSV2dhfr; lane 2, control cell line C, a clone amplified to 0.64 μM MTX resistance from control cell line A; lane 3, clone 1, an unamplified primary clone cotransfected with pSV2-dec and pSV2dhfr; lane 4, clone 31, a clone amplified to 0.32 μM MTX resistance from clone 1; lane 5, clone 61, a clone amplified to 0.64 μM MTX resistance from clone 1. See Table 1 for the details of each clone.

Methods. The expression vector, pSV2-dec, was constructed by ligating the 1.8 kilobase (kb) full-length decorin cDNA into the 3.4-kb *Hind*III-*Bam*HI fragment of pSV2 (ref. 10). Dihydrofolate reductase (dhfr)-negative CHO cells (CHO-DG44, kindly provided by Dr Lawrence A. Chasin) were cotransfected with pSV2-dec (25 μg) and pSV2dhfr (1 μg) by the calcium phosphate coprecipitation method. The transfected cells were selected in nucleoside-free α -modified minimal essential medium (GIBCO) supplemented with 9% dialysed fetal calf serum, 2 mM glutamine, 100 units ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. Colonies arising from transfected cells were picked using cloning cylinders, expanded and checked for the expression of decorin by immunoprecipitation of $^{35}\text{SO}_4$ -labelled culture supernatants and competitive enzyme-linked immunosorbent assay (ELISA, see Table 1). Clones expressing detectable amounts of decorin were then subjected to gene amplification by culture in the presence of increasing concentrations of MTX¹². The amplified cell lines were cloned either by limiting dilution or by picking a single MTX-resistant colony. Stock cultures of these established cell lines were propagated in MTX-containing medium. Before use in experiments, the cells were subcultured in MTX-free medium from stock cultures and passed at least once in this medium to eliminate possible MTX effects. Controls were transfected only with pSV2dhfr and treated exactly as the experimental cells thereafter. A total of 17 decorin-expressing and control cell lines were isolated. Metabolic labelling of the cells with $^{35}\text{SO}_4$ or [^3H]leucine and immunoprecipitation was performed as described previously⁶.

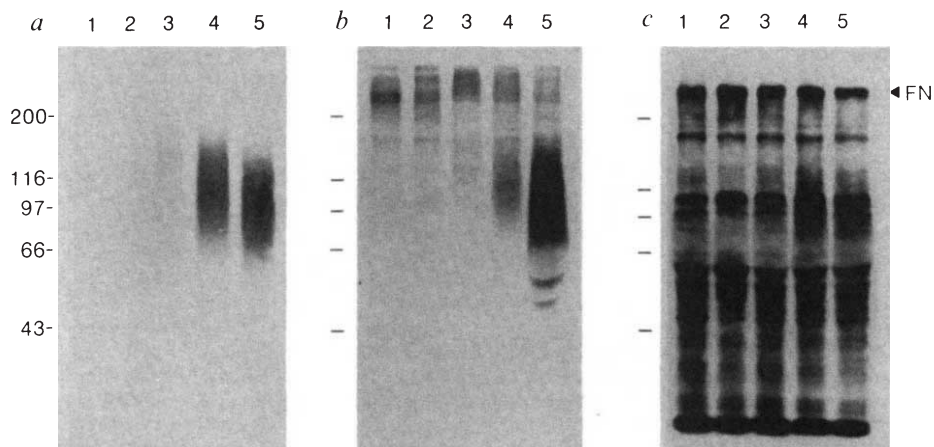
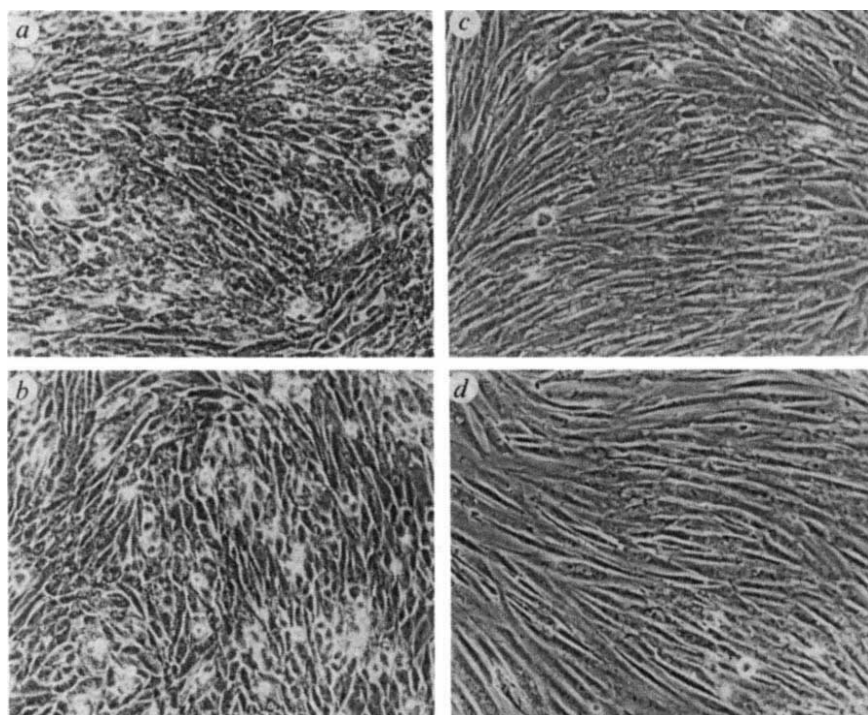


Fig. 2 Morphological changes accompanying the expression of decorin in CHO cells. *a*, Control cell line A, an unamplified line. *b*, Control cell line C, amplified to 0.64 μM MTX resistance. *c*, Clone 31, a medium level expressor, amplified to 0.32 μM MTX resistance. *d*, Clone 61, a high level expressor, amplified to 0.64 μM MTX resistance. All cell lines were seeded and grown to confluency in MTX-free culture medium.



has been suggested for cell-surface heparin sulphate proteoglycans²⁰⁻²². We have found recently (Y.Y. and E.R., unpublished data) that the decorin-expressing cells deposit more fibronectin into their extracellular matrix than the control cells. Decorin may, therefore, exert its effect by promoting fibronectin deposition. Finally, because decorin is one of the two proteoglycans whose synthesis is upregulated by transforming growth

factor β (ref. 23), our results raise the possibility that the inhibition of cell proliferation caused by this growth factor in some cells²⁴ could be directly related to the increased production of decorin. Taken together, our results suggest that contact-mediated inhibition of cell proliferation may be mediated through production and deposition of extracellular matrix components.

Table 1 Decorin production and spreading on substrate of transfected cells

Clone	Transfection	MTX resistance (μM)	Decorin production ($\text{pg cell}^{-1} \text{ day}^{-1}$)*	Area of spreading ($\mu\text{m}^2 \text{ cell}^{-1}$)†
Control cell line A	pSV2dhfr	0	0	2,725 $\pm$ 627
Control cell line B	pSV2dhfr	0.32	0	2,585 $\pm$ 693
Control cell line C	pSV2dhfr	0.64	0	2,659 $\pm$ 586
Clone 1	pSV2-dec + pSV2dhfr	0	0.2	3,087 $\pm$ 842
Clone 31	pSV2-dec + pSV2dhfr	0.32	4	4,759 $\pm$ 898
Clone 33	pSV2-dec + pSV2dhfr	0.32	11	5,554 $\pm$ 1,002
Clone 66	pSV2-dec + pSV2dhfr	0.64	14	5,482 $\pm$ 1,382
Clone 61	pSV2-dec + pSV2dhfr	0.64	25	6,472 $\pm$ 1,147

* Cells were seeded in 24-well plates in MTX-free culture medium at a density of 10^5 cells per well. After 24 hours, medium was replaced with fresh medium (0.3 ml per well) and the cells were incubated another 24 hours. Concentration of decorin in this culture supernatant was determined by competitive ELISA. Briefly, a mixture of culture supernatant and rabbit antipeptide antibody against decorin was incubated in the wells of microtitre plates coated with decorin which had been purified from human fetal membranes⁶. The amount of antibody bound to the wells was determined by alkaline-phosphatase-conjugated goat antirabbit IgG as a second antibody and the concentration of decorin was determined by using purified decorin as a standard. The cells were counted by haemocytometer at the end of the 24-hour incubation.

† Cells (1.2×10^5) were seeded in a 60-mm culture dish in MTX-free culture medium. After 6 hours, the cultures were fixed with 3% paraformaldehyde and stained with toluidine blue. Quantitative evaluation of spreading was performed by measuring the surface area of the cells with a surface integration program of an image analyser (Cue-2, Olympus). The mean and standard deviation of the values from 50 cells are shown for each cell line.

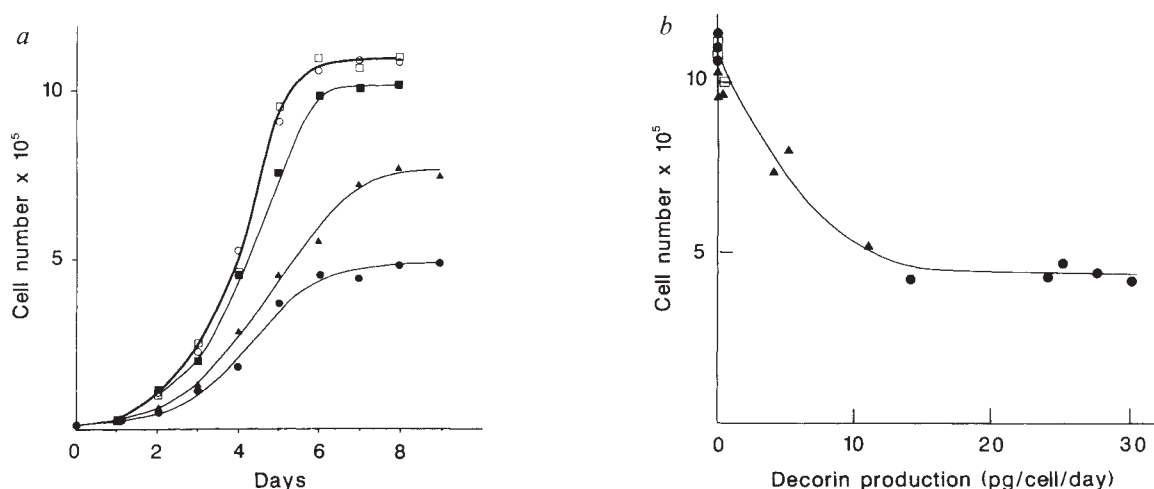


Fig. 3 Growth rates and saturation densities of decorin-expressing and control CHO cells in culture. *a*, Growth curves of five representative cell lines expressing various levels of decorin. (□) Control cell line A; (○) control cell line C; (■) clone 1; (▲) clone 31; (●) clone 61. *b*, Correlation of expression level of decorin and saturation densities for 17 decorin-expressing and control cell lines. Abscissa: decorin production per cell per day as determined by ELISA. Ordinate: the saturation cell density in a 16-mm well of 24-well plate. (●) Cell lines amplified to 0.64 μM MTX resistance; (▲) cell lines amplified to 0.32 μM MTX resistance; (□) unamplified cell lines.

Methods. For growth curves, cells were plated in triplicate at a density of 2×10^4 cells per well in 24-well plates and cultured in 1 ml of MTX-free culture medium. Half of the medium was replaced with fresh medium every 24 hours. At regular intervals, cells in some of the wells were collected by trypsinization and counted by haemocytometer. The means of the triplicate were shown. The saturation densities were determined from the growth curves as a cell number at the plateau of the curve. Decorin production was determined as in Table 1.

We thank Dr Tom Krusius for his help at the early stages of this work and Drs Mario Bourdon, Eva Engvall and Robert Oshima for comments on the manuscript. This work was supported by grants from the National Cancer Institute and a postdoctoral fellowship to Y.Y. from the Cancer Research Institute.

Received 3 August; accepted 6 October 1988.

1. Ruoslahti, E. & Pierschbacher, M. D. *Science* **238**, 491-497 (1987).
2. Gospodarowicz, D., Ferrara, N., Schweigerer, L. & Neufeld, G. *Endocr. Rev.* **8**, 95-114 (1987).
3. Ruoslahti, E. *A. Rev. Cell Biol.* (in the press).
4. Pearson, C. H., Winterbottom, N., Fackre, D. S., Scott, P. G. & Carpenter, M. R. *J. biol. Chem.* **258**, 15101-15104 (1983).
5. Glössi, J., Beck, M. & Kresse, H. *J. biol. Chem.* **259**, 14144-14150 (1980).
6. Brennan, M. J., Oldberg, Å., Pierschbacher, M. D. & Ruoslahti, E. *J. biol. Chem.* **259**, 13742-13750 (1984).

7. Krusius, T. & Ruoslahti, E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7683-7687 (1986).
8. Day, A. A., McQuillan, C. I., Termine, J. D. & Young, M. R. *Biochem. J.* **248**, 801-805 (1987).
9. Schmidt, G. *et al. J. Cell Biol.* **104**, 1683-1691 (1987).
10. Mulligan, R. C. & Berg, P. *Science* **209**, 1423-1427 (1980).
11. Subramani, S., Mulligan, R. & Berg, P. *Molec. Cell Biol.* **1**, 854-864 (1981).
12. Kaufman, R. J. & Sharp, P. A. *J. molec. Biol.* **159**, 601-621 (1982).
13. Knox, P. & Wells, P. *J. Cell Science* **40**, 77-88 (1979).
14. Brennan, M. J., Oldberg, Å., Hayman, E. G. & Ruoslahti, E. *Cancer Res.* **43**, 4302-4307 (1983).
15. Rosenberg, L. C., Choi, H. U., Poole, A. R., Lewandowska, K. & Culp, L. A. *Ciba Fdn Symp.* **124**, 47-68 (1986).
16. Fritze, L. M. S., Reilly, C. F. & Rosenberg, R. D. *J. Cell Biol.* **100**, 1041-1049 (1985).
17. Ishihara, M., Fedarko, N. S. & Conrad, H. E. *J. biol. Chem.* **262**, 4708-4716 (1987).
18. Vogel, K. G. & Heinegård, D. *Biochem. J.* **223**, 587-597 (1984).
19. Scott, J. E. *Biochem. J.* **218**, 229-233 (1984).
20. Saunders, S. & Bernfield, M. *J. Cell Biol.* **106**, 423-430 (1988).
21. Izzard, C. S., Radinsky, R. & Culp, L. A. *Expl. Cell Res.* **165**, 320-336 (1986).
22. Woods, A., Couchman, J. R., Johansson, S. & Höök, M. *EMBO J.* **5**, 665-670 (1986).
23. Bassols, A. & Massagué, J. *J. biol. Chem.* **263**, 3039-3045 (1988).
24. Sporn, M. B., Roberts, A. B., Wakefield, L. M. & de Crombrughe, B. *J. Cell Biol.* **105**, 1039-1045 (1987).

Ion channel of acetylcholine receptor reconstructed from images of postsynaptic membranes

Chikashi Toyoshima & Nigel Unwin

Medical Research Council Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK

The nicotinic acetylcholine receptor belongs to a class of molecules that respond transiently to chemical stimuli by opening a water-filled channel through the cell membrane for cations to diffuse. This channel lies along the central axis delineated by a ring of five homologous, membrane-spanning subunits^{1,2} and thus has properties, such as conductance and ion selectivity, which depend on the profile created by the encircling subunits. Insight has been gained recently about the amino-acid residues implicated directly in the ion transport³⁻⁷, and some information about the subunit configuration around the channel has come from electron microscopy studies of postsynaptic membranes crystallized in the form of flattened tubular vesicles⁸⁻¹¹. The resolution along the axis of the channel has, however, been limited by the restricted range of views obtainable. Here we report the structure of the channel at 17 Å resolution, determined by three-dimensional image reconstruction from tubular vesicles having receptors organized in helical arrays across their surfaces. The helical symmetry is preserved by suspending the tubes in thin films of ice, and the receptors in such tubes can be seen from all angles, allowing the channel to be revealed clearly in relation to the lipid bilayer and the peripheral protein for the first time.

The long and narrow tubes were obtained for this study (Fig. 1a) from postsynaptic membranes of the electric ray, *Torpedo marmorata*, and consist of dimers of receptor molecules periodically arranged in the native lipids⁸. Before investigating the helical structure, we first examined features in thin sections (Fig. 1b). The tubes, like isolated fragments of postsynaptic membranes^{12,13}, contain peripheral proteins, seen as a uniformly thick layer decorating their inner surface. These proteins prob-

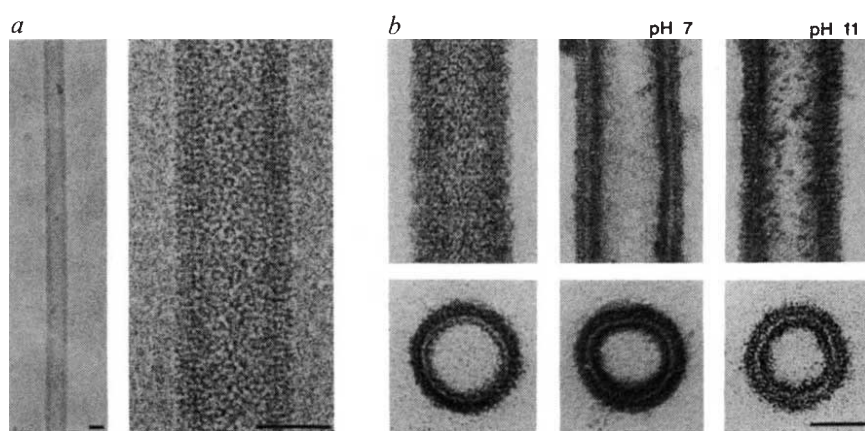
ably consist predominantly of the protein of relative molecular mass 43,000 (43K) (refs 14-16), which associates weakly with the receptor and is released from the membranes (unless treated with *N*-ethylmaleimide¹⁷) by alkaline pH¹⁸. The proteins coat more than 50% of the inner surfaces of the tubes, assuming that the very thin layer of stain sometimes observed beneath the membrane (Fig. 1b, lower left) corresponds only to the receptor. Moreover, we find (Fig. 1b, right) that the proteins redistribute in an irregular way on the surface after short exposure to alkaline pH. Because of *N*-ethylmaleimide treatment^{10,17}, they are not released and the ordered arrangement of the receptors is not significantly disturbed. Thus the tubes have different cytoplasmic structures at different pHs, providing a means to distinguish the receptor and peripheral protein components from one another.

For the reconstructions we recorded images of the frozen tubes (Fig. 1a) in weakly and then strongly defocused pairs¹⁹. Diffraction patterns of these images (Fig. 2a) show a series of layer lines characteristic of a helical structure and extending to a resolution of ~16 Å. Three-dimensional maps were calculated from layer-line data summed from both images (Fig. 2b), and represent information averaged over about 4,000 molecules.

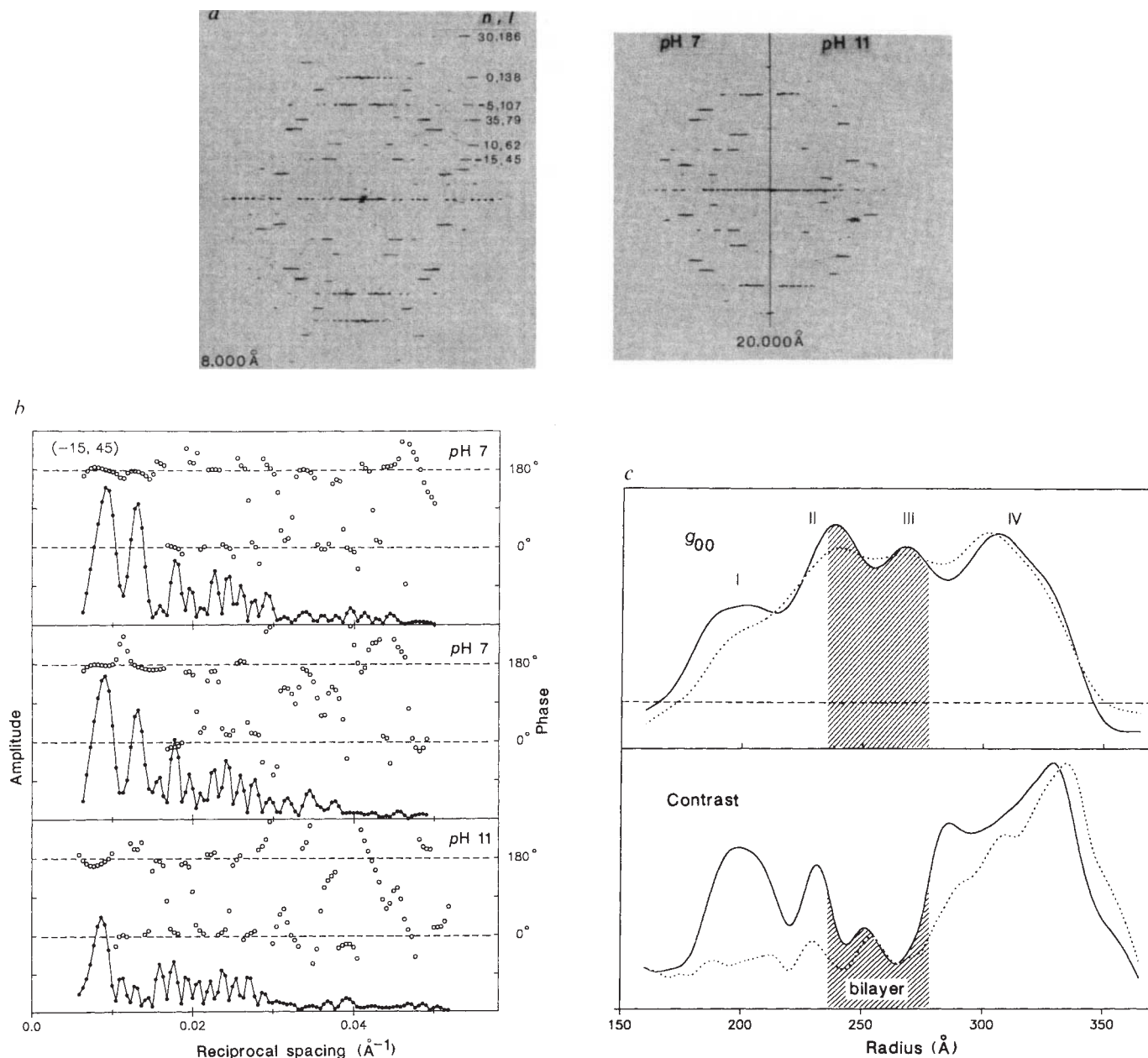
Figure 3a shows a section normal to the tube axis in the three-dimensional map obtained for the native tubes. The receptors form an oblique lattice on the cylindrical surface so that this section intercepts several of them in different places. Two circumferential weakly contrasted bands are also present; the walls of the channel made by the synaptic portion of the receptor project ~65 Å from the outermost band, whereas cytoplasmic densities extend ~55 Å from the innermost band.

The two bands must be a feature of the lipid bilayer. First, they are at the same radii in relation to the rest of the structure, as the edges of the membrane seen clearly in the thin sections (Fig. 3a, insert). Second, their separation (30 Å) is identical to that measured from images of receptor-containing vesicles, also present in the same preparation (results not shown). Third, the bands are at the same radii as those regions of the structure where the contrast between the protein and the outside environment is weakest (Fig. 2c). This finding is anticipated because the density of the protein should be matched by the phospholipid headgroups and steroid moiety of the cholesterol in the outer

Fig. 1 Images of tubes suspended in ice (a), and embedded in plastic and sectioned (b). The sections are of the tubes cut longitudinally (upper row) and transversely (lower row), the membrane appearing as a thin white line. The longitudinal sections are of the tube surface (left), showing the good preservation of receptors, and of regions just below the surface before (middle) and after (right) exposure to alkaline pH. The transverse sections are of the native (left, middle) and alkali-treated (right) tubes. The protein coat, indicated by the uniformly thick layer of stain lining the inner membrane surface, is sometimes missing (lower left), despite the well preserved structure outside the membrane. This coat, where present, is disrupted and reorganized in an irregular way by the alkaline pH (right); as the receptors retain their position in the crystal lattice, the coat must represent non-receptor protein. Scale bar, 500 Å.



Methods. Tubes for the helical analysis, as in a, were maintained in 100 mM sodium cacodylate, pH 6.8, or dialysed against 100 mM sodium phosphate, pH 11.0, immediately before freezing. The freezing was performed by plunging holey carbon grids, to which the specimens had been applied, into liquid ethane slush. Electron microscopy was at 120 kV, using a Philips EM400T or EM420 equipped with a low-dose kit and Gatan cryo-holder, maintained at -168 °C. Images were recorded of tubes over holes in the carbon support film (magnification, $\times 36,000$; dose, ~ 8 electrons Å⁻²), at $\sim 8,000$ Å then $\sim 20,000$ Å underfocus¹⁹ using Kodak SO163 film. Tube diameters had to be < 750 Å to maintain helical symmetry. For the sectioned specimens, suspensions of the tubes in 100 mM sodium cacodylate were dialysed against sodium phosphate buffer either at pH 7.0 or pH 11.0 (4 °C, 40 min) and subsequently against fixative: 0.8% tannic acid (Mallenkrodt), 4.0% glutaraldehyde and 100 mM sodium phosphate, pH 7.0 (1 h); fixative was then mixed directly (1:1) with the suspension (30 min); specimens were pelleted, washed, reacted with 1% osmium tetroxide and embedded in the usual way; sections were post-stained with uranyl acetate and lead citrate.



part of the bilayer²⁰ more nearly than by the low-density hydrocarbon chains in the inner part or by the surrounding ice.

The outer and inner leaflets of a lipid bilayer have not previously been resolved in electron image reconstructions, and the peak-to-peak separation of 30 Å is 10–15 Å smaller than might have been anticipated on the basis of profile X-ray diffraction measurements from typical biological membranes²⁰ or details seen in cryo-images of pure phospholipid vesicles²¹. Two factors could contribute to the smaller spacing observed. First, the scattering properties of electrons are different from those of X-rays: the oxygen and phosphorus atoms in the headgroup regions will be less emphasized by the electrons as the electron-scattering amplitudes for these atoms, relative to carbon atoms, are smaller than for X-rays²². Second, the postsynaptic membrane lipids contain a high proportion of cholesterol (almost 50 mol%²³) which is intercalated between the phospholipid fatty acid chains. The denser steroid portions of these molecules are about 30 Å apart and contribute strongly to the X-ray diffraction profile²⁰. Therefore, the appearance of the postsynaptic membrane in the electron images is probably dominated by the

cholesterol, rather than by the phospholipids. To check on this interpretation we prepared artificial vesicles of dipalmitoyl phosphatidylcholine (the major phospholipid component of the postsynaptic membrane²³) and demonstrated that the peak-to-peak separation in the cholesterol-free bilayer is much larger (>40 Å; data not shown).

Figure 3b shows another view of the native tube, from the inside, made by stacking cylindrical sections at successive radii on top of one another. At the outermost radii (most distant from the observer), the densities corresponding to the five receptor subunits all begin at about the same level, symmetrically encircling the ~25-Å-wide synaptic entrance to the channel. They follow paths that are almost parallel to the axis of the molecule throughout most of their length, passing through the lipid bilayer and emerging as five symmetrically disposed densities on the cytoplasmic side. These densities project 15–20 Å from the estimated position of the phospholipid headgroups. At smaller radii, the densities are replaced by a single, central density. The change in structure, from approximately pentagonal to a single density, is abrupt, as reflected by the two well separated peaks in the contrast plot (Fig. 2c).

◀ **Fig. 2** Diffraction patterns computed from images of the ice-embedded tubes (a), plots of amplitude and phase variations along a typical layer-line (b), and variations in radial density and contrast (c). The upper diffraction pattern is of a tube imaged at 8,000 Å underfocus (the first image of the pair; the indices n and l refer to the Bessel orders and layer-line numbers); the lower diffraction pattern (20,000 Å underfocus) is a composite comparing data from the native (left) and alkali-treated tubes (right); the layer-lines are not at identical levels on the two sides because of small differences in the unit cell of the surface lattice. The layer-line plots (b) show the amplitude and phase variations as a function of reciprocal spacing along the $(-15, 45)$ layer-line; the three examples compare data used for the native map (average of ~4,000 molecules; upper), a second set of unused data (average of ~1,500 molecules; middle) and data for the high pH map (average of ~4,000 molecules; lower); this layer line corresponds to the $(1, 0)$ reflection in analyses of the flattened tubes⁸⁻¹¹; the transforms have been shifted to a common twofold origin, so that the phases would ideally be 0 or 180°; note that the first 10 major peaks in the upper two plots all have similar amplitudes and identical phases, whereas the equivalent peaks in the plot from the alkali-treated tubes show substantial differences; for example, the second peak (reciprocal spacing, 0.013 Å⁻¹) has a much weaker amplitude and an opposite phase to the corresponding peak from the native tubes. The variations in mean radial density g_{00} (c, upper) and contrast (c, lower) are across tubes frozen at pH 7 (full lines) and pH 11 (dotted lines). In the density plots (c, upper), peaks I and IV correspond to structures on the cytoplasmic and synaptic sides of the lipid bilayer respectively; peaks II and III correspond to the circumferential bands (Fig. 3a, c) associated with the bilayer; these curves, calculated from the composite equatorial layer-line data, relate closely to the edge profile in the images (as in Fig. 1a for example) seen by viewing the tubes at a glancing angle along their axes; the peak positions were shown not to be affected by the defocus level or by contrast transfer function correction. The contrast plots (c, lower) represent the differences between minimum and maximum densities at various levels throughout the two three-dimensional maps. The regions of weakest contrast, in the bilayer-spanning portion of the structure, are at the same radii as peaks II and III above, indicating that the protein and the bilayer in these regions have almost matching densities; the extent of the bilayer to the estimated positions of the phospholipid headgroups is indicated by the shaded area. On the cytoplasmic side of the bilayer (radius, 190–245 Å) the native structure shows two regions in which features are well contrasted, one (radius, 190–225 Å) associated with the peripheral protein, and the other (radius, 225–245 Å) associated with the part of the receptor projecting from the phospholipid headgroups; in the alkali-treated structure the contrast associated with the peripheral protein has disappeared, indicating a loss of regularity.

Methods. Appropriate areas were densitometered (scanning aperture and step size, 15 µm), converting the images into two-dimensional arrays of optical densities. Helical reconstruction methods, as outlined below, were different in several ways from those commonly used²⁹. The image arrays were interpolated so that layer lines in the Fourier transforms were in exact register with the sampling grid to provide accurate Fourier terms; for the interpolation, the repeat distance and the rotation of the tube relative to the array were measured by cross-correlating two different parts along the tube. Layer lines in Fourier transforms calculated from 512 × 1,024 image arrays (typically 4,000 Å long portions of the tube) were extracted and the two sides averaged to improve the signal-to-noise ratio. Further improvement in signal-to-noise ratio was achieved on the basis of the presence of twofold symmetry normal to the helix axis, and amplitude-weighted average phase deviations from the ideal values (0° or 180°) were used as a measure of the quality of the images. Refinement of the twofold phase origin, and the fitting of two images at different defoci and of several stretches along the same tube were done in two steps. First, only strong layer lines were used to put different transforms to a common twofold origin. Then all the layer lines allowed by the selection rule were extracted to the cut-off resolution imposed by the defocus (~16 Å and ~30 Å for the weakly and strongly defocused images, respectively) and averaged. The phase residual for twofold symmetry was calculated for each layer line, which was discarded if the residual was >45°. These procedures allowed us to extract as much information as possible and provided significant improvement in the quality of the data: phase residuals for the twofold symmetry were 12–18° (strongly defocused image), 24–30° (weakly defocused image) before averaging, and were 15.6° (native; 36 layer-lines) and 19.9° (alkali-treated; 27 layer-lines) for the final composite datasets. Corrections were not made for the effects of non-uniform contrast transfer except along the equator, because the range in contrast transfer for the composite off-equatorial Fourier terms was small (0.4 for the lowest resolution terms, to 1.0); equatorial data were corrected assuming 7% amplitude contrast¹⁹.

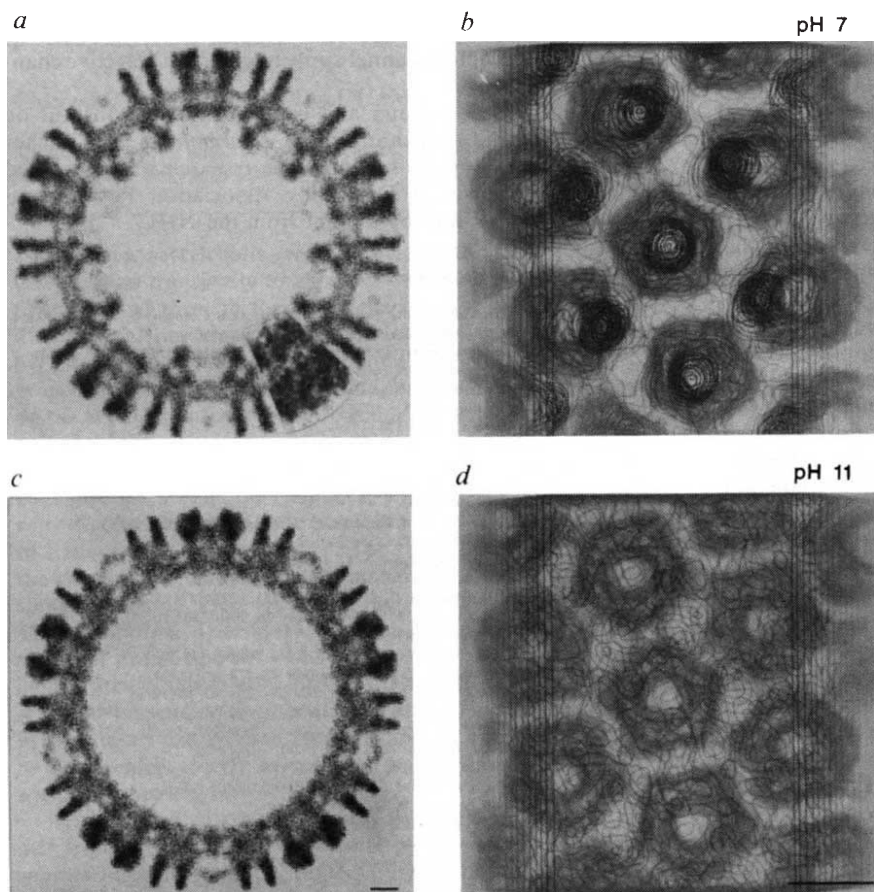


Fig. 3 Three-dimensional maps obtained by helical image reconstruction of the native tubes (a, b) and tubes at pH 11 (c, d); (a) and (c) are sections normal to the tube axis; (b) and (d) are views from the tube axis, looking outwards, made by stacking cylindrical sections at successive radii (spacing corresponding to 10 Å). The section (a) shows the lipid bilayer, the synaptic portion of the receptor projecting outwards, and the cytoplasmic portion and peripheral protein on the inside; a sector from a transverse thin section (Fig. 1b) is inserted to show the good correspondence in the position of the lipid bilayer visualized by the two methods; the outer diameter of the tube is ~700 Å. The view (b) shows the peripheral protein overlying the pentagonally symmetrical receptor at the cytoplasmic entrance, which projects just above the bilayer surface; the outermost contours enclose a total volume of ~310,000 Å³ for the receptor and ~35,000 Å³ for the peripheral protein; the volume for the receptor is about 85% of that expected for the glycosylated protein³⁰. c, A section equivalent to that in a for the alkali-treated tubes (except for very slight differences in the units cells of the surface lattices); the lipid bilayer is not as clear, appearing to be slightly disrupted by the alkaline pH (ref. 31), and the densities lining the inner surface have changed. d, View equivalent to that in b, for the alkali-treated tubes showing the pentagonal outline made by the receptor subunits, but not the overlying peripheral protein. Scale bar, 50 Å.

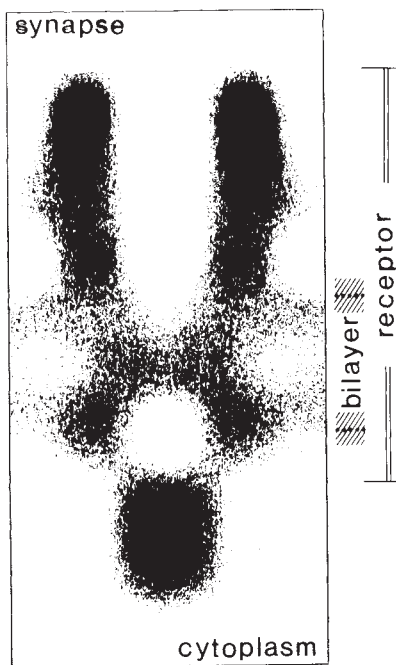


Fig. 4 Axial section through the cylindrically averaged native structure, showing details of the ion channel in relation to the lipid bilayer (estimated positions of the phospholipid headgroups indicated) and peripheral protein (bottom). The cylindrical averaging was done about the pseudo-fivefold axis of a single molecule.

Next, we analysed images of tubes frozen in alkaline solution, conditions which cause the peripheral protein to redistribute in an irregular way over the cytoplasmic surface (Fig. 1b) while the receptor retains its position in the crystal lattice. Diffraction patterns of these images were different, with marked changes along some of the layer lines (Fig. 2a and b). The three-dimensional map (Fig. 3c and d) reflected these changes: over the synaptic and bilayer-spanning portions, the structure was similar to the native one (Fig. 3a), but at the cytoplasmic end the central density, prominent in the native structure (Fig. 3b), was completely missing (Fig. 3d). Hence the central density, occupying a volume corresponding to about 30K, must be due to the peripheral protein. The location of the peripheral protein at the extremity of the receptor correlates well with the thin-section results (Fig. 1b) suggesting that the receptor projects only a small amount from the bilayer on the cytoplasmic side.

Although we have not directly identified the peripheral protein, it is reasonable to suppose that it corresponds to the 43K protein. The 43K protein is the only subsynaptic protein known to be present in concentrations similar to the receptor¹⁵; furthermore, morphological studies combined with antibody labelling have shown that the 43K protein does contribute extra density at the cytoplasmic end²⁴. The redistribution over the membrane surface induced by the alkaline pH suggests that it could also be tethered to the lipid^{16,25}. Various functions have been attributed to the 43K protein (refs 16 and 26); our findings lend support to a possible regulatory role.

The ion channel has several distinctive features, which are best illustrated with the structure obtained by cylindrically averaging the densities composing a single molecule about its pseudo-fivefold axis. Figure 4 is an axial section through the native structure. It shows the long vertical wall made by the encircling receptor subunits, the pair of transverse densities made by the lipid bilayer, and matter at the bottom of the receptor, identified as a peripheral protein rather than as a component of the receptor itself. This protein overlies the axial

path of the channel, apparently touching some of the receptor subunits.

The ion channel at the synaptic end is about 25 Å across. It extends downwards towards the bilayer, changing little in cross-section over the next 60 Å. Then, close to the level of the outer phospholipid headgroups and over a very short distance, it becomes too narrow to be resolved (probably <10 Å diameter). The channel is unresolved across the interior of the bilayer, but at about the level of the inner phospholipid headgroups, it widens again, and continues with a wide (~20 Å diameter) cross-section to the cytoplasmic end, a further 15–20 Å away. The channel is partitioned quite sharply at the level of the phospholipid headgroups and may be considered to have a tripartite structure, consisting of two entrance domains and the 'pore' in between. We suggest that the three domains, and the short junctional regions, have specific roles in facilitating the flow of cations across the membrane. The entrance domains are tubular rather than funnel-shaped; their wall linings have a radius (10–13 Å) large enough to allow the ions to diffuse freely, but small enough in relation to the Debye distance²⁷ for surface charges to affect ion concentrations and thereby provide more efficient ion selectivity²⁸. The junctions between the entrances and the pore form 'platforms', and charged residues located at these levels may act more directly in defining cation selectivity and the efficiency of transport. The pore itself is not resolved, but must have a much smaller radius than the entrance domains; presumably, it prevents the passage of large ions when agonist binds to induce the open state, and blocks ion flux altogether when the agonist is released.

Recent experiments, involving covalent labelling of the receptor with non-competitive channel blockers^{3,4} and conductance measurements on chimaeric molecules⁶, imply that at least a part of the putative α -helix M2 lines the pore. Details in the three-dimensional maps are in good accord with this possibility, as the pore appears to be about 30 Å long, roughly the length of the M2 helix; the charged amino acids that flank this helix could well be at the platforms bracketing the pore. It will be interesting to see if these structural characteristics of the acetylcholine receptor channel apply to other ion-selective channels as well.

We thank Dr C. Cazaux and colleagues at the Institut de Biologie Marine, Arcachon for *T. marmorata* and Nichol Thomson for cutting the sections. The research was supported in part by a Muscular Dystrophy Association Postdoctoral Fellowship to C.T. and by a grant from the NIH.

Received 15 August; accepted 4 October 1988.

- Karlin, A. in *The Cell Surface and Neuronal Function* (eds Cotman, C. W., Poste, G. & Nicolson, G. L.) 191–260 (North-Holland, Amsterdam, 1980).
- Changeux, J.-P., Devillers-Thiery, A. & Chemouille, P. *Science* **225**, 1335–1345 (1984).
- Giraudat, J. et al. *Biochemistry* **26**, 2410–2418 (1987).
- Hucho, F. W., Oberthur, W. & Lottspeich, F. *FEBS Lett.* **205**, 137–142 (1986).
- Sakmann, B. et al. *Nature* **318**, 538–543 (1985).
- Imoto, K. et al. *Nature* **324**, 670–674 (1986).
- Mishina, M. et al. *Nature* **313**, 364–369 (1985).
- Brisson, A. & Unwin, P. N. T. *J. Cell Biol.* **99**, 1202–1211 (1984).
- Brisson, A. & Unwin, P. N. T. *Nature* **315**, 474–477 (1985).
- Kubalek, E. et al. *J. Cell Biol.* **105**, 9–18 (1987).
- Unwin, P. N. T., Toyoshima, C. & Kubalek, E. *J. Cell Biol.* **107**, 1123–1138 (1988).
- Sealock, R. *J. Cell Biol.* **92**, 514–522 (1982).
- Sealock, R., Wray, B. E. & Froehner, S. C. *J. Cell Biol.* **98**, 2239–2244 (1984).
- Sobel, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **75**, 510–514 (1978).
- LaRochelle, W. J. & Froehner, S. C. *J. biol. Chem.* **261**, 5271–5274 (1986).
- Carr, C., McCourt, D. & Cohen, J. B. *Biochemistry* **26**, 7090–7102 (1987).
- Barrantes, F. J. *J. Cell Biol.* **92**, 60–68 (1982).
- Neubig, R. R. et al. *Proc. natn. Acad. Sci. U.S.A.* **76**, 690–694 (1979).
- Toyoshima, C. & Unwin, P. N. T. *Ultramicroscopy* **25**, 279–292 (1988).
- Caspar, D. L. D. & Kirschner, D. A. *Nature new Biol.* **231**, 46–52 (1971).
- Lepault, J., Pattus, F. & Martin, N. *Biochim. biophys. Acta* **820**, 315–318 (1985).
- International Tables for X-ray Crystallography* Vol. 3 (Reidel, Holland, 1983).
- Popot, J.-L. et al. *Eur. J. Biochem.* **85**, 27–42 (1978).
- Popot, J.-L. et al. *J. Cell Biol.* **105**, 1829–1846 (1987).
- Porter, S. & Froehner, S. C. *J. biol. Chem.* **258**, 10034–10040 (1983).
- Froehner, S. C. *Trends Neurosci.* **9**, 37–41 (1986).
- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, 1984).
- Dani, J. A. & Eisenman, G. *J. gen. Physiol.* **89**, 959–983 (1987).
- DeRosier, D. J. & Moore, P. B. *J. molec. Biol.* **52**, 355–369 (1970).
- Popot, J.-L. & Changeux, J.-P. *Physiol. Rev.* **64**, 1162–1239 (1984).
- Bonini, I. C. et al. *Biochem. J.* **245**, 111–118 (1987).

Movement and segregation of kinetochores experimentally detached from mammalian chromosomes

B. R. Brinkley*, R. P. Zinkowski*, W. L. Mollon*,
F. M. Davis†, M. A. Pisegna†, M. Pershouse‡
& P. N. Rao†

* Department of Cell Biology and Anatomy, University of Alabama at Birmingham, UAB Station, Birmingham, Alabama 35294, USA

† Department of Medical Oncology and ‡ Department of Hematology, University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA

The kinetochore is a specialized structure at the centromere of eukaryotic chromosomes that attaches chromosomes to the mitotic spindle^{1,2}. Recently, several lines of evidence have suggested that kinetochores may have more than a passive role in the movement of chromosomes during mitosis and meiosis. Kinetochores seem to attract and 'capture' microtubules that grow from the spindle poles and microtubules may lengthen or shorten by the addition or subtraction of tubulin subunits at their kinetochore-associated ends³⁻⁶. An attractive hypothesis is that kinetochores function as 'self-contained engines running on a microtubule track'⁵. Here, we show that kinetochores can be experimentally detached from chromosomes when caffeine is applied to Chinese hamster ovary cells that are arrested in the G1/S phase of the cell cycle. The detached kinetochore fragments can still interact with spindle microtubules and complete all the mitotic movements in the absence of other chromosomal components. As these cells enter mitosis before DNA synthesis is completed, chromosome replication need not be a prerequisite for the pairing, alignment and segregation of kinetochores.

When Chinese hamster cells were arrested at the G1/S boundary of the cell cycle by treatment with 2 mM hydroxyurea for 20 h and subsequently treated for 4–6 h with caffeine (5 mM), the cells entered mitosis without completing DNA synthesis as originally observed in BHK (baby hamster kidney) cells⁷. We called these cells 'mitotic cells with unreplicated genome' (MUGs) and followed the assembly of the mitotic apparatus and related structures in these cells by indirect immunofluorescence and electron microscopy.

MUGs were observed in two Chinese hamster cell lines—Chinese hamster ovary (CHO) and V79-8—within 2 h following the addition of caffeine (Fig. 1a), and a mitotic index of 76% was achieved within four hours of treatment with both caffeine and Colcemid (demecolcine). Flow cytometer measurements of DNA in mitotic cells indicated that most MUG cells contained a 2C level of DNA whereas normal Colcemid-arrested mitotic cells had the expected 4C content (Fig. 1b). These results confirm that caffeine can induce S-phase arrested CHO cells to enter mitosis without completion of DNA synthesis.

When control CHO cells were arrested in mitosis with Colcemid and stained by indirect immunofluorescence using kinetochore-specific human autoantibodies from sera of scleroderma 'CREST' (calcinosis, Raynaud's phenomenon, oesophageal dysmotility, sclerodactyly, telangiectasia) patients, the kinetochores appear as a pair of bright fluorescent spots located on each side of the centromere (Fig. 2a). When MUGs were stained by the same procedure, the chromosomes were highly fragmented and dispersed throughout the cell. The entire complement of kinetochores were displaced from the chromosomes and grouped in a cluster near the cell centre (Fig. 2b). Confirmation that the kinetochores were indeed detached from the chromosomes was obtained when MUGs were examined by transmission electron microscopy (Fig. 2c–e).

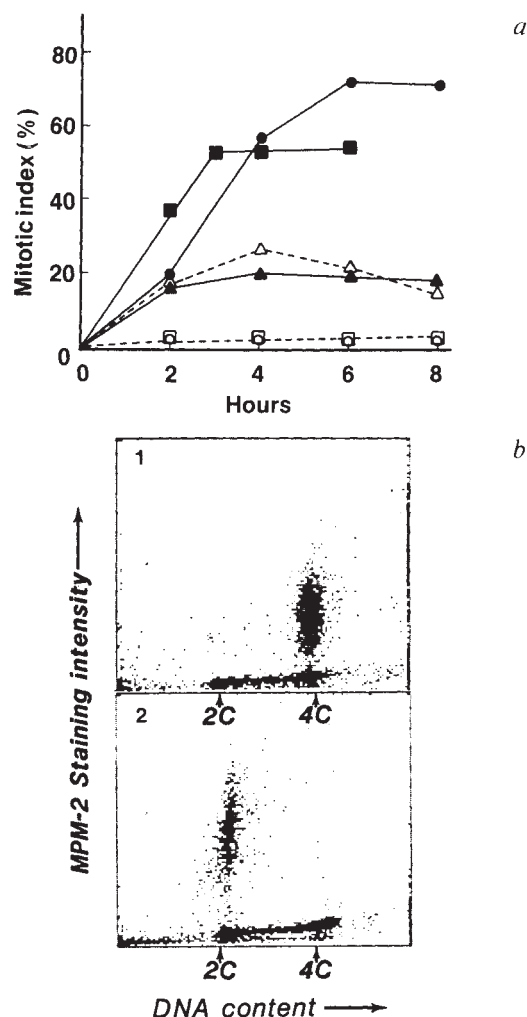


Fig. 1 *a*, Kinetics of mitotic accumulation during caffeine treatment. Mitotic CHO cells were plated at a density of 10^6 cells per 28 cm^2 in a number of dishes and incubated with hydroxyurea (2 mM) for 20 h. They were treated in different ways (at $t=0$), and incubation continued for 8 h with cell samples being taken at 2-h intervals. The mitotic index is plotted as a function of time after the addition of caffeine. The treatments were: Colcemid ($0.25\text{ }\mu\text{g ml}^{-1}$), ○; caffeine (5 mM), ●; Colcemid + caffeine, ▲; Colcemid + caffeine + actinomycin D ($5\text{ }\mu\text{g ml}^{-1}$), ■; and Colcemid + caffeine + cycloheximide ($25\text{ }\mu\text{g ml}^{-1}$), □ to cultures blocked in early S phase by hydroxyurea. Under identical experimental conditions (that is, hydroxyurea + caffeine + Colcemid), the rate of mitotic accumulation in V79-8 cells, ■, was significantly greater than that of CHO cells. *b*, Correlated flow cytometric analysis of DNA content and MPM-2 (a mitosis-specific monoclonal antibody specific for non-histone phosphoprotein¹⁰) staining of MUG cells. CHO mitotic cells and MUGs obtained by selective detachment were mixed with equal numbers of random, unsynchronized CHO cells and doubly stained for MPM-2 antibody reactivity and DNA content. CHO mitotic cells (1) stained by MPM-2 have a 4C DNA content. CHO MUG cells (2) stained by MPM-2 have a 2C DNA content.

Individual kinetochores with varying amounts of associated chromatin and with their characteristic trilaminar plate-like ultrastructure were seen tethered to spindle microtubules in the cytoplasm (Fig. 2e). In most profiles, the kinetochores appeared as ovoid structures with an electron-dense chromatin core surrounded by the kinetochore plates. Microtubules were often seen attached to the outermost surface which also had a prominent 'fuzz' or corona (Fig. 2d–e). In fortuitous sections, it was possible to identify kinetochores which seemed to be undergoing detachment from the centromere at the time of fixation (Fig. 2d). If MUGs were induced in the presence of Colcemid or

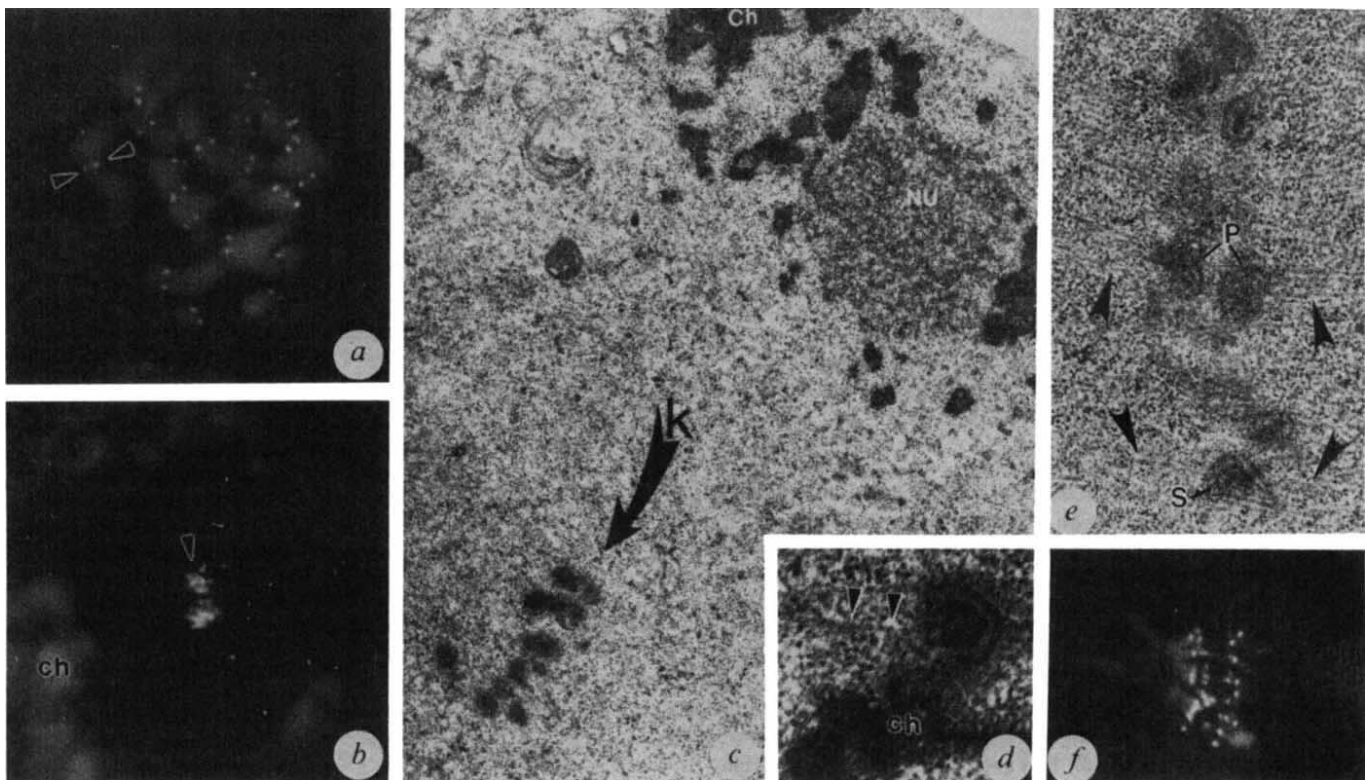


Fig. 2 Normal and detached kinetochores. *a*, Normal chromosomes of CHO cells. Kinetochores stained with scleroderma CREST auto-antibodies appear as paired fluorescent spots located on opposite sides of the centromere of each chromosome (arrows). *b*, Caffeine-treated CHO cell showing entire complement of kinetochores (arrow) detached from surrounding fragmented chromosomes (ch) and aligned near the cell centre. *c*, Electron micrograph showing kinetochores (arrow, k) segregated from chromosomes (ch) and nucleolus (NU). *d*, Higher magnification showing kinetochores during detachment from chromosome (ch). Note microtubule (arrows) attached to kinetochore plate. *e*, Micrograph showing the alignment of paired (P) and unpaired (S) kinetochores on the metaphase plate. The arrows point to microtubules extending from the kinetochores toward opposite poles. *f*, Segregation of detached kinetochores at the start of anaphase as shown by indirect immunofluorescence staining. These cells were arrested before the completion of DNA synthesis, and therefore pairing, alignment and segregation do not require centromeric replication.

nocodazole, spindle microtubules failed to form and kinetochores remained associated with the condensed chromosome fragments (Fig. 3*a-c*). Therefore, microtubule attachment to the kinetochores seems to be essential for pulling kinetochores away from the chromosomes.

The assembly of the mitotic spindle and progression of cells through mitosis was observed in synchronized populations of cells by inducing MUGs in the presence of Colcemid and then washing out the microtubule inhibitor. By collecting cells at brief intervals during recovery from Colcemid arrest, and double-staining them with anti-tubulin and anti-kinetochore antibodies, it was possible to analyse spindle assembly and kinetochore movement in many cells synchronized at various stages of mitosis. The sequence of photomicrographs in Fig. 3 demonstrates clearly that during spindle microtubule assembly, kinetochores become detached from the fragmented chromosomes, migrate to the metaphase plate and undergo anaphase movement to the poles. Subsequently, cytokinesis and daughter cell formation occurs. Without kinetochores, the residual chromosome fragments become randomly distributed to daughter cells (Fig. 3*l, o*) forming numerous micronuclei.

To analyse the capacity of detached kinetochores to undergo successful anaphase segregation in MUGs, we counted the number of stained kinetochores segregated to pairs of daughter cells at late telophase (Fig. 4). The number of kinetochores per pair of daughter cells ranged from 14 to 64 for MUGs and 40 to 50 (modal number 44 corresponding to chromosome number of $2N=22$) for normal CHO cells. MUGs with near 'diploid' kinetochore numbers segregated more efficiently than those with fewer kinetochores.

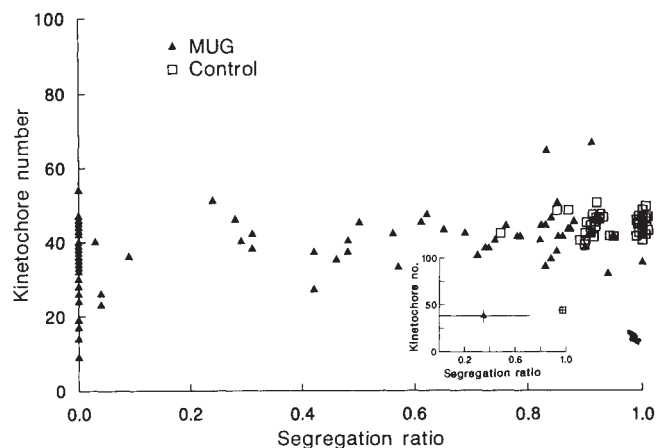
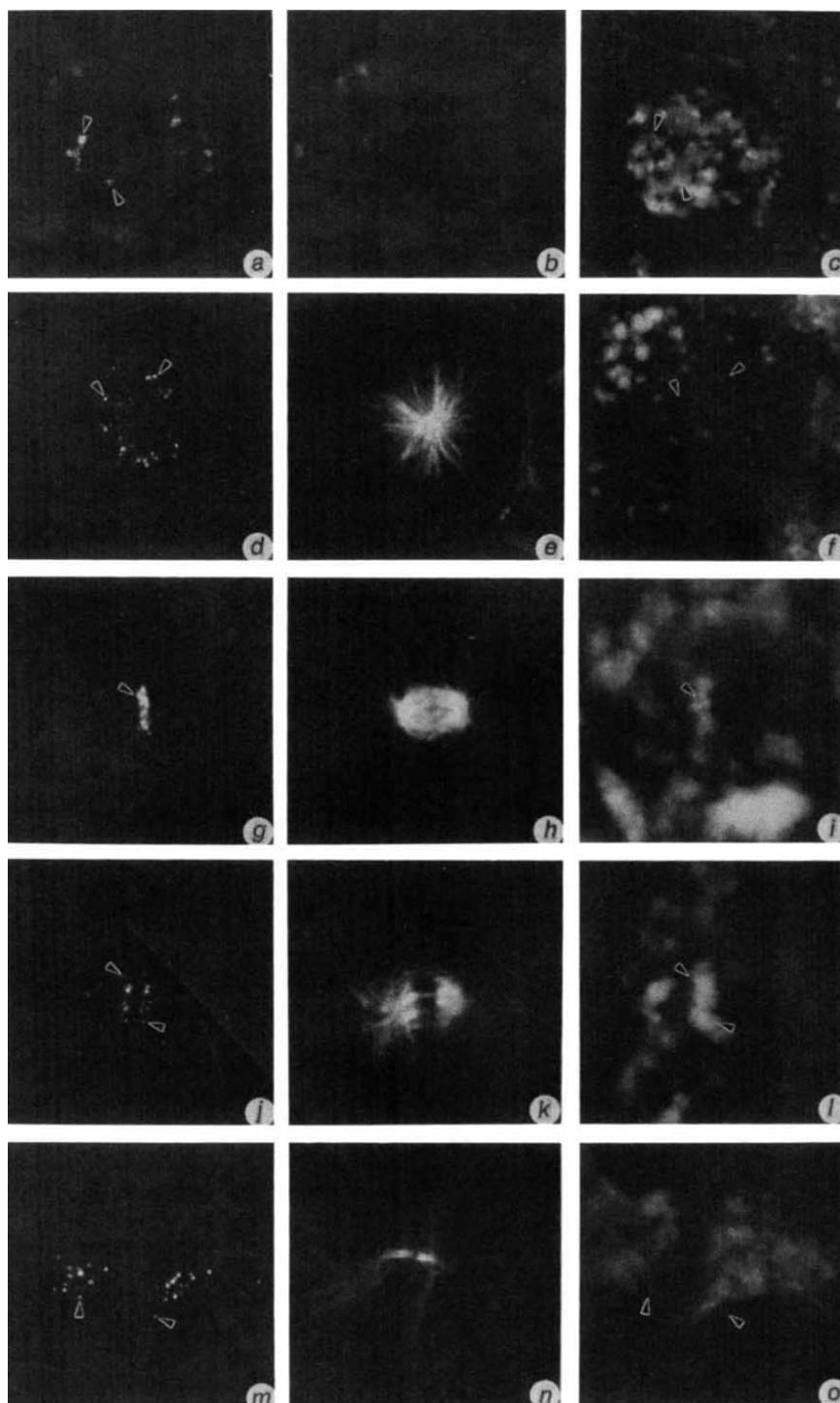


Fig. 4 The number of kinetochores in pairs of daughter cells (total number of fluorescent spots as in Fig. 3*m*) and their fidelity of segregation as measured by the segregation ratio was obtained by dividing the number of kinetochores in one daughter cell by the number in the other. Thus, a ratio of 1.0 indicates equal segregation at anaphase, whereas a ratio of 0 indicates a total lack of segregation. Normal CHO cells contain a modal diploid chromosome number of $2N=22$ and a total of 44 kinetochores. As expected these cells display a high segregation fidelity. MUGs displayed a wide range of kinetochore numbers and segregation fidelities. Only those MUGs with a near diploid number of kinetochores segregate with high fidelity. The inset in the lower right-hand section of this figure presents a plot of the standard deviation of values for kinetochore number and segregation ratio in control (□) and MUGs (▲).

Fig. 3 Movement, alignment and segregation of detached kinetochores in MUGs. Cells incubated for 20 h in 2 mM hydroxyurea were synchronized in mitosis by treatment with 5 mM caffeine for 6 h and the addition of Colcemid ($0.1 \mu\text{g ml}^{-1}$) for the last 3 h. The Colcemid was washed out allowing spindle formation and progression through mitosis. Cells were stained by indirect immunofluorescence with fluorescein anti-kinetochore (*a, d, g, j, m*) and rhodamine anti-tubulin antibodies (*b, e, h, k, n*) and with Hoescht 33258 for DNA (*c, f, i, l, o*). The position of the kinetochores with respect to the spindle and fragmented chromosomes can be seen during Colcemid arrest (*a, b, c*) and in each stage of mitosis beginning with a monopolar spindle (*d, e, f*) and in progressing to metaphase (*g, h, i*), anaphase (*j, k, l*) and telophase (*m, n, o*). For reference, the arrows point to positions of selected kinetochores in each set of cells.



In normal mitosis, the migration of chromosomes to the metaphase plate and their proper alignment and equational segregation is thought to be facilitated by paired kinetochores positioned on opposite faces of the chromatids^{8,9}. Because MUGs have a 2C DNA content (Fig. 1b) it is reasonable to conclude that their chromosomes are single chromatids with unreplicated kinetochores. The normal segregation unit consists of a pair of sister chromatids in mitotic cells with a 4C DNA content, so how can segregation occur in cells with 2C DNA? There are at least three possible explanations: (1) the DNA of the centromere/kinetochore region can undergo replication even though the rest of the genome does not; (2) centromeric DNA replication is inhibited but single unreplicated kinetochores can associate in pairs, migrate to the metaphase plate and become

segregated at anaphase; or (3) single unreplicated kinetochores can interact with microtubules extending from opposite poles and become positioned on the metaphase plate; segregation can occur by either splitting the kinetochore into two halves or by random migration to one pole or the other.

We were unable to detect any evidence of centromeric DNA replication by inducing MUGs in the presence of ^3H -thymidine and then processing the cells for combined immunofluorescence and *in situ* autoradiography. Therefore, the alignment and segregation of kinetochores in MUGs seems to occur in the absence of kinetochore replication. We propose that the near diploid number of fluorescent spots seen in some cells is not due to replication but to splitting or fragmentation of the detached kinetochores by action of the drugs or from forces applied by

the mitotic spindle. Careful analysis of MUGs at anaphase by immunofluorescence and electron microscopy provides strong evidence that the kinetochore fragments can align on the metaphase plate either as single units attached to microtubules arising from opposite poles or as paired units facing opposite poles (Fig. 2e). Obviously, some combinations of these arrangements are favourable for segregation at anaphase. Those cells with the correct number of kinetochore fragments arranged in favourable orientations segregate normally whereas those with too few and with unfavourable orientations either fail to segregate or do so unevenly.

These experiments demonstrate that fully differentiated kinetochores of mammalian chromosomes can be experimentally uncoupled from their chromosomal components but still retain the capacity to interact with spindle microtubules and undergo the full repertoire of mitotic movements. Cells with the correct number and orientation of kinetochores display normal patterns of mitotic segregation. As kinetochore replication is apparently blocked in MUGs, splitting or fragmentation of the kinetochores may account for the near diploid number of fluorescent spots observed in some cells. Contained within

the kinetochore fragments and their associated spindle microtubules are the molecular components needed for pairing, congression to the metaphase plate and anaphase segregation.

We thank Ann Harrell for typing the manuscript and R. Balczon, P. Gregory and A. Tousson for proofreading and helpful discussions. We also thank Claude Bennett, William Koopman and Graciela Alarcon for providing the scleroderma CREST serum and L. I. Binder for microtubule monoclonal antibodies. This work was supported in part by HHS/NCI grants to B.R.B. and P.N.R.

Received 16 May; accepted 4 October 1988.

1. Brinkley, B. R., Tousson, A., Valdivia, M. M. in *Aneuploidy, Etiology and Mechanisms* (eds Dellarco, V. L., Voytek, P. E. & Hollaender, A.) 243-368 (Plenum, New York, 1985).
2. Rieder, C. L. *Int. Rev. Cytol.* **79**, 1-58 (1982).
3. Pickett-Heaps, J. D., Tippit, D. H. & Porter, K. R. *Cell* **79**, 729-744 (1982).
4. Mitchison, T. L., Evans, L., Schulze, E. & Kirschner, M. *Cell* **45**, 515-527 (1986).
5. Gorbisky, G. L., Sammak, P. J. & Borisy, G. G. *J. Cell Biol.* **104**, 9-18 (1987).
6. Gorbisky, G. L., Sammak, P. J. & Borisy, G. G. *J. Cell Biol.* **106**, 1185-1192 (1988).
7. Schlegel, R. & Pardee, A. B. *Science* **232**, 1264-1266 (1986).
8. Nicklas, R. B. in *Aneuploidy, Etiology and Mechanisms* (eds Dellarco, V. L., Voytek, P. E. & Hollaender, A.) 183-195 (Plenum, New York, 1985).
9. Ostergren, G. *Hereditas* **37**, 85-156 (1951).
10. Davis, F. M., Tsao, T. Y., Fowler, S. K. & Rao, P. N. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2926-2930 (1983).

Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein

E. S. Bochkareva, N. M. Lissin & A. S. Girshovich*

Institute of Protein Research, Academy of Sciences of the USSR, Pushchino, Moscow Region, USSR

It has been suggested that newly synthesized proteins are maintained in their unfolded state by cellular ATP-driven factors which may prevent or reverse the formation of misfolded structures or promote the correct assembly of oligomeric proteins or post-translational secretion¹⁻⁶. Using a photocross-linking approach, we have identified the 20S heat-shock GroEL protein as the major cytosolic component which forms a complex with the unfolded newly synthesized pre- β -lactamase or chloramphenicol acetyltransferase in *Escherichia coli*. Dissociation of these complexes is ATP-dependent. The unfolded state of pre- β -lactamase, maintained by the transient interaction with GroEL, may be essential for the secretion of this protein.

The cross-linking method we used is a modification of our procedure⁷ in which initiator N-acyl-Met-tRNA^{Met} is used as a photolabel donor because of its suitability for prokaryotic translation systems. The N-acyl is the 4-(3-trifluoromethyl-diazirine) benzoyl residue⁸, a precursor of the highly reactive carbene radical that can be photoactivated. The newly synthesized protein can carry the photolabel at one site only, at its N-terminal methionine. For translation, we used the S30 extract of *E. coli* MRE600 and the transcript of plasmid pH624 (ref. 9) which encodes pre- β -lactamase (pre- β -lac). The photoactive group and irradiation were both needed for the post-translational cross-linking of pre- β -lac to the S30 components, leading to the appearance of some radioactive bands above the pre- β -lac, one with a major band of relative molecular mass (M_r) of about 120,000 (120K) (Fig. 1a, lanes 1-3). The irradiated mixture was ultracentrifuged and, surprisingly, the 120K product sedimented at about ~20S (Fig. 2a), indicating that the pre- β -lac cross-links to a component of the 20S particle. The increase in intensity of the pre- β -lac band itself in this region confirms that this interaction does occur. The major protein sedimenting like

the 120K product is the GroEL protein (see arrow in Fig. 2c) which has M_r of about 65K and exists as a 20-25S tetradecamer^{10,11}. To test whether GroEL is cross-linked to pre- β -lac, we isolated the GroEL particles (Fig. 2d, lane 2 and electron micrographs). Using antibodies raised against GroEL (specificity shown in Fig. 2d, lanes 3 and 4) and immobilized on BrCN-Sepharose, we obtained an S30(-GroEL) preparation that had retained translational activity but had lost 90-95% of its GroEL. The centrifugation experiment was repeated using this S30(-GroEL) preparation (Fig. 2b). As we expected, there was no increase in intensity of the pre- β -lac band in the 20S region in the absence of GroEL, indicating that pre- β -lac is indeed part of the complex. Figure 1 shows that GroEL removal leads to the disappearance of the 120K band (lane 6); it appears again with the addition of GroEL (lanes 7 and 8). Moreover, anti-GroEL specifically binds the 120K product (lane 9). Thus, pre- β -lac is indeed cross-linked with GroEL.

We performed an analogous series of experiments with the pK13 plasmid encoding chloramphenicol acetyltransferase (CAT). Here also, GroEL was the main acceptor of the newly synthesized protein (Fig. 1b): GroEL addition led to the appearance of one major band of M_r ~110K (lanes 10 and 11) that was recognized by anti-GroEL (lane 13). In the presence of membranes the cross-linking is sharply lowered for the secretory protein (pre- β -lac, compare lanes 3 and 5), but not for the cytosolic protein (CAT; compare lanes 11 and 12). Although only two proteins were tested, we assume that other proteins could also interact with GroEL.

The crucial question is whether it is the unfolded or folded state of the newly synthesized protein that interacts with GroEL. The stimulation of pre- β -lac cross-linking with GroEL at increased dithiothreitol (DTT) concentration (Fig. 1, compare lanes 3 and 4) indicates that the protein may be unfolded because thiols impede the formation of S-S bonds which stabilize the folded state (pre- β -lac can form S-S bonds *in vitro*¹²) and because increased DTT concentration increases the efficiency of pre- β -lac post-translational secretion (Fig. 3c, lanes 11 and 12) which is known¹³⁻¹⁷ to require the secreting proteins to be unfolded. Further evidence for the affinity of GroEL for the unfolded pre- β -lac is shown in Fig. 3a: increase in the time between translation and GroEL addition leads to a gradual fall in the 120K band intensity. This is not caused by pre- β -lac proteolysis as the total amount of pre- β -lac is preserved during incubation (Fig. 3c, lanes 1-3). It is therefore reasonable to explain the decreased cross-linking, that is, the decrease in the affinity of GroEL for pre- β -lac, as a consequence of pre- β -lac

* To whom correspondence should be addressed.

Fig. 1 Photoactivated cross-linking of newly synthesized pre- β -lac (a) or CAT (b) with components of the *E. coli* S30 extract.

Methods. Transcript preparation (for 5 μ l): 0.5 μ g pH624 (ref. 9) or pKC13 (constructed from pKK223-3 and pCML, Pharmacia) plasmids, 0.5 μ g *E. coli* RNA polymerase, 7.5 nmol of each NTP, 10 mM magnesium acetate, 150 mM potassium acetate, 0.2 mM EDTA, 0.1 mM DTT and 40 mM triethanolamine acetate, pH 7.9 were incubated for 30 min at 37°C. Preparation of the initiation complex (for 15 μ l): 5 μ l transcript, 2 μ l S30 (about 35 μ g total protein, 0.9 pmol ribosomes) or S30(-GroEL), 1.5 pmol N-acyl-[¹⁴C]Met-tRNA^{Met} (0.2 Ci mmol⁻¹), 4.5 mM ATP, 0.9 mM GTP were incubated for 5 min at 37°C and cooled in ice. Translation (for 24 μ l): 15 μ l initiation complex, 2.4 μ g total tRNA, 12 μ Ci [³⁵S]Met (300 Ci mmol⁻¹) were included with other components as previously described²²; DTT concentration was 20 mM, except in sample 4 in which it was 1 mM. GroEL was added to samples (0.17, 0.5, 0.35 and 0.35 μ g to samples 7, 8, 11 and 12, respectively) which were then incubated for 15 min at 37°C, and cooled in ice. 10 mM cold methionine (1.2 μ l) was added and the mixture irradiated for 3 min by a high-pressure mercury lamp SVD-120A (USSR) with a filter transmitting light above 310 nm. In samples 5 and 12, membranes (up to 3 absorbance units at 280 nm) were added and incubated for 10 min at 25°C before irradiation. Samples were then precipitated with ice-cold trichloroacetic acid (TCA) and either analysed by 10–18% SDS-polyacrylamide gel electrophoresis (PAGE) according to Laemmli (lanes 10–12) or treated first with affinity-purified antibodies (48 μ l translation mixtures): 1 μ g anti- β -lac (lanes 1–8) or 10 μ g anti-GroEL (lanes 9 and 13). After incubation for 10 min at 37°C and an overnight incubation at 4°C, the antigen-antibody complexes were isolated on protein A-Sepharose (ref. 7) and analysed by SDS-PAGE (the bands corresponding to the cross-linking products shown in lanes 9 or 13 were absent if pre-treatment with anti-GroEL was omitted). Kodak X-ray DEF-2 film was used for autoradiography. [¹⁴C]-labelled size markers are shown at the left. S30 was prepared by lysozyme lysis of cells followed by treatment with *Staphylococcus aureus* nuclease²³. S30(-GroEL) was obtained by passing S30 through anti-GroEL-Sepharose. Membranes (inverted membrane vesicles) were isolated as described²². It is unclear why the apparent M_r values, deduced from the electrophoretic mobility of cross-linked products (for pre- β -lac or CAT), exceed the sum of their components.

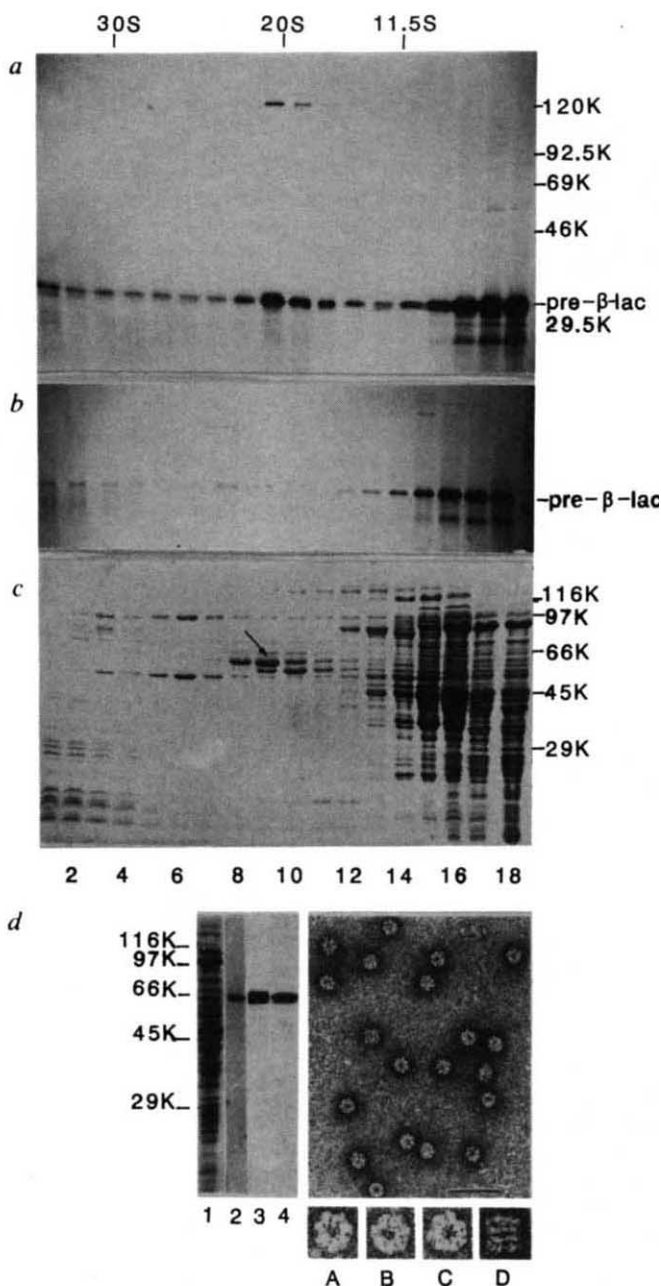
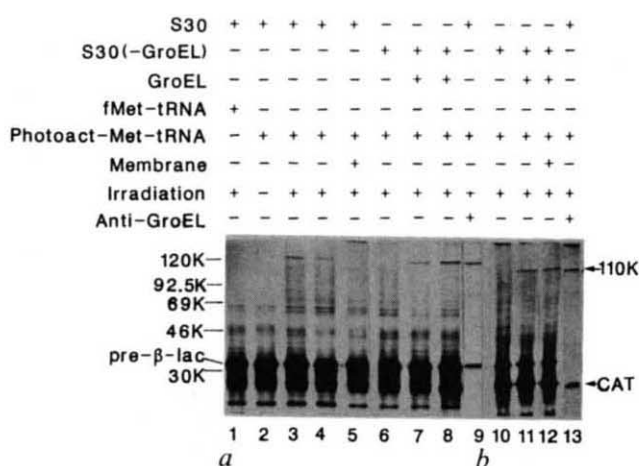


Fig. 2 Ultracentrifugation of S30 (a) and S30(-GroEL) (b) mixtures after pre- β -lac synthesis and irradiation, staining of S30 proteins by Coomassie blue (c) and characterization of purified GroEL and anti-GroEL antibodies (d).

Methods. a and b, Potassium acetate (up to 0.5 M) was added to 24 μ l irradiated mixture (about 0.3 pmol [³⁵S]-labelled pre- β -lac) which was layered on to a 10–25% sucrose gradient in 0.5 M potassium acetate, 5 mM magnesium acetate, 0.1 mM EDTA, 5 mM DTT and 40 mM triethanolamine-acetate, pH 7.5, and spun at 33,000 r.p.m. for 15 h at 4°C in an SW50 rotor. The fractions were then precipitated with TCA, analysed by SDS-PAGE (8–18%) and autoradiographed. The positions of the 30S ribosomal subunits and 11.5S bovine liver catalase are marked. c, S30 (20 μ l) was layered on a parallel sucrose gradient and stained with Coomassie blue after centrifugation and SDS-PAGE. Size markers are shown on the right. d, Lane 1, S30 total protein; lane 2, GroEL; lanes 3 and 4, immunoblotting of GroEL and S30 total protein respectively with anti-GroEL. Electron micrographs of the purified GroEL, negatively stained with uranyl acetate are shown at the right^{10,11}: top, general field; bottom, selected electron micrographs in the top (A–C) or side (D) projections. Bar is 500 Å. GroEL was isolated from S30 by ultracentrifugation in conditions similar to those described in a. Fractions containing the GroEL were recentrifuged and chromatographed on Mono-Q (Pharmacia) in 20 mM Tris-HCl, pH 7.5, 5 mM mercaptoethanol on a gradient linear in NaCl. GroEL eluted at 0.43 M NaCl and was analysed by SDS-PAGE (8–20%). Anti-GroEL antibodies were prepared by immunization of rabbits with purified GroEL and serum chromatography on GroEL-Sepharose. Antibody specificity was tested by immunoblotting followed by peroxidase staining. Size markers are shown at the left.

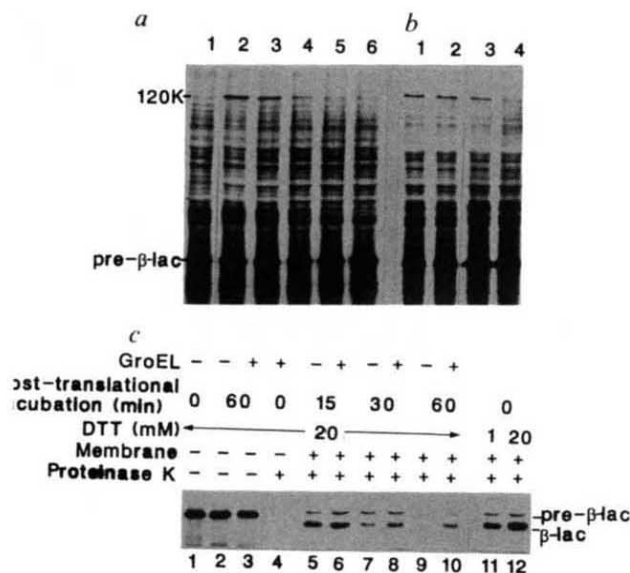


Fig. 3 GroEL interacts with the unfolded form of pre-β-lac. *a*, Effect of post-translational incubation of pre-β-lac on the yield of the 120K product. *b*, Effect of protein addition on the yield of the 120K product. Myoglobin, apomyoglobin and heat-denatured apomyoglobin are shown in lanes 2, 3 and 4 respectively. Lane 1, control. *c*, Effect of GroEL on post-translational translocation of pre-β-lac.

Methods. *a*, Cold methionine (1.2 μl; 10 mM) and chloramphenicol (up to 0.1 mg ml⁻¹) were added to 24 μl mixture containing S30(-GroEL) after translation. GroEL (0.35 μg) was added either before translation (lane 2) or immediately after chloramphenicol addition (lane 3), or after 15, 30 and 60 min of additional incubation at 37 °C (lanes 4, 5 and 6, respectively). The mixtures were then incubated for 10 min at 25 °C, irradiated in ice, precipitated with TCA and analysed by SDS-PAGE (10–18%). Sample 1 does not contain GroEL. *b*, After translation was stopped with chloramphenicol, 0.2 μg GroEL and 5 μg sperm whale myoglobin, apomyoglobin or heat-denatured apomyoglobin were added to the 24 μl mixtures containing S30(-GroEL) (lanes 2, 3 or 4, respectively), incubated for 10 min at 25 °C and irradiated in ice. Sample 1 contained 0.2 μg GroEL only. *c*, The translational mixtures (13 μl) contained fMet-tRNA^{Met} and S30(-GroEL) (lanes 1–10) or S30 (lanes 11–12). GroEL (0.27 μg) was added before translation. Translation was stopped with chloramphenicol (0 min) and inverted membrane vesicles (0.03 absorbance units at 280 nm) and NADH (up to 2 mM) were added after 15, 30 or 60 min of post-translational incubation at 37 °C. After translocation (20 min, 25 °C), 0.3 μg proteinase K was added and after a further 20 min at 25 °C, proteolysis was blocked with phenylmethylsulphonylfluoride (up to 2 mM for 30 min at 25 °C). The samples were then precipitated with TCA and analysed by SDS-PAGE. Samples 1–3 contained 4 μl translation mixture. A similar result to that shown in lane 4 was obtained by proteolysis of the sample with membranes and 2% *n*-octylglucoside (not shown). The presence of a small band of pre-β-lac in lanes 5–12 does not depend on the incompletely digested untranslocated protein (compare with control in lane 4) and probably represents translocated but unprocessed chains²².

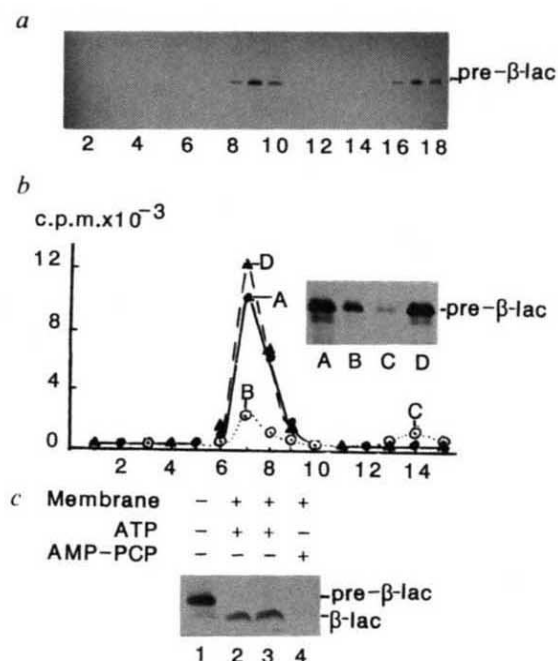


Fig. 4 Isolation and stability of the GroEL/pre-β-lac complex. *a*, Ultracentrifugation of S30 mixture after pre-β-lac synthesis. *b*, Effect of ATP or AMP-PCP on the stability of the isolated complex and *c*, on the translocation competence of the pre-β-lac in the isolated complex.

Methods. *a*, Translational mixture (150 μl, about 1.3 pmol ³⁵S-labelled pre-β-lac) was ultracentrifuged, as described in Fig. 2a legend. fMet-tRNA was used to initiate translation. GroEL (3 μg; about 4 pmol of 20S GroEL particles) was added before translation to increase the yield of complex. 10% of each fraction was precipitated with TCA, analysed by SDS-PAGE and autoradiographed. Fraction 9 from the sucrose gradient (0.3 ml), with most of the 20S complex, was diluted with five volumes of the centrifugation buffer without sucrose and potassium acetate, and concentrated into a volume of 0.3 ml at 4 °C on a Centricon 10 (Amicon). *b*, The isolated complex was incubated for 10 min at 25 °C with 5 mM ATP (○ · · · ○) or AMP-PCP (▲ · · · ▲) or on its own (● — ●; control) and recentrifuged in the buffer containing 0.5 mM ATP or AMP-PCP or no additions respectively. One-third of each 0.36-ml gradient fraction was then counted in the Triton X100/toluene scintillator (the same picture was obtained if the samples were boiled with 5% TCA and the precipitates filtered through Millipore filters and counted; data not shown). A–D on the graph indicate the fractions analysed by SDS-PAGE (see insert). The main peak of radioactivity corresponding to the dissociated pre-β-lac was not present, presumably because of its high adsorption on the tube surface. ATP treatment does not change the position of GroEL itself on the sucrose gradient (data not shown). *c*, Inverted membrane vesicles (0.1 absorbance unit at 280 nm) and NADH (up to 2 mM) were added to the isolated complex (28 μl, about 0.02 pmol ³⁵S-labelled pre-β-lac) and incubated for 20 min at 25 °C in the presence of 4 mM ATP (lanes 2 and 3) or AMP-PCP (lane 4). One sample (lane 3) was pre-incubated for 30 min at 37 °C without ATP and vesicles. Further treatment with proteinase K followed by SDS-PAGE was as described in Fig. 3c legend. Lane 1 (control) corresponds to the isolated complex (28 μl).

folding during pre-incubation in the absence of GroEL. In addition, the GroEL/pre-β-lac interaction is inhibited by myoglobin, in proportion to the extent to which the added protein is unfolded (Fig. 3b): the addition of myoglobin does not alter the yield of the 120K product (lane 2); both the removal of heme (substitution of apomyoglobin for myoglobin), leading to the partial unfolding of the added protein¹⁸ (lane 3), and the addition of heat-denatured apomyoglobin (lane 4) significantly

lower the cross-linking efficiency. These results indicate that GroEL interacts with the unfolded form of newly synthesized proteins.

Because secretory proteins are post-translationally translocated across the membranes in the unfolded state, the question arises of how the interaction between GroEL and the unfolded pre-β-lac affects its secretion. The secretion of pre-β-lac was almost completely blocked after about 30 min (at 37 °C) post-

translational pre-incubation without GroEL (Fig. 3c, lanes 7 and 9) but was still active if pre-incubated in the presence of GroEL (lanes 8 and 10). We infer that this is because GroEL is maintaining an unfolded state of pre- β -lac that is competent for secretion.

To study the properties of the complex containing GroEL and newly synthesized proteins, we isolated it by ultracentrifugation (Fig. 4a, fraction 9). As purified GroEL has an ATPase activity¹¹, we tested the effect of ATP on the stability of the isolated complex (Fig. 4b): GroEL/pre- β -lac complex dissociates with ATP but not with its non-hydrolysable analogue adenylyl(β - γ -methylene)-diphosphonate (AMP-PCP). The same results were obtained for CAT (data not shown). When the isolated GroEL/pre- β -lac complex was added to the membrane preparation, pre- β -lac translocated in the presence of ATP, but not in the presence of AMP-PCP (Fig. 4c, lanes 2 and 4), namely in conditions under which the pre-formed complex would dissociate. Pre-incubation of this complex for 30 min at 37 °C without ATP does not decrease subsequent ATP-dependent translocation of pre- β -lac (Fig. 4c, lanes 2 and 3). Thus, the interaction of GroEL with unfolded proteins is abolished by ATP hydrolysis. Moreover, the unfolded state of the secretory protein, maintained by the transient interaction with GroEL, is competent for secretion of the protein and this is consistent with the effect of GroEL on the crude translation-secretion system (Fig. 3c). GroEL (the 20S particle) is similar in this respect to the trigger-factor ($M_r \sim 60K$) (ref. 16) or the *secB* gene product ($M_r \sim 12K$) (ref. 19), which promote protein secretion in *E. coli* by preventing the folding of secretory proteins as this makes them translocation-incompetent. The ATP-dependent release of the unfolded secretory protein from its complex with GroEL may be a component of the activating effect that ATP has generally on the post-translational protein translocation⁵.

The properties of GroEL that we have described here are consistent with the suggestion¹ that heat-shock proteins, of which GroEL is one²⁰, have an affinity for unfolded or misfolded proteins and impede aggregation (or promote disaggregation) of the latter by using the energy of ATP hydrolysis. Moreover, the GroEL particle is a stimulator of protein assembly, being involved transiently in the initial steps of the head or tail formation of many coliphages^{20,21}. We propose that the transient ATP-driven interaction of the GroEL particle with newly synthesized proteins before they fold finally can affect whether their fate lies in folding, assembly or in secretion pathways.

We thank Professor A. S. Spirin for helpful discussion, Dr V. D. Vasiliev for electron microscopy of the GroEL particles and Drs N. I. Matvienko and O. B. Yarchuk for the plasmids pHc624 and pKc13.

Received 15 July; accepted 20 September 1988.

- Pelham, H. R. B. *Cell* **46**, 959-961 (1986).
- Rothman, J. E. & Kornberg, R. D. *Nature* **322**, 209-210 (1986).
- Zimmermann, R. & Meyer, D. I. *Trends biochem. Sci.* **11**, 512-515 (1986).
- Ellis, J. *Nature* **328**, 378-379 (1987).
- Hurt, E. C. *Trends biochem. Sci.* **12**, 369-370 (1987).
- Pelham, H. R. B. *Nature* **332**, 776-777 (1988).
- Kurzchalia, T. V. *et al.* *Nature* **320**, 634-636 (1986).
- Nassal, M. *Justus Liebig's Annl. Chem.* **9**, 1510-1523 (1983).
- Boros, J., Postfai, G. & Venetianer, P. *Gene* **30**, 257-260 (1984).
- Hohn, T., Hohn, B., Engel, A., Wurtz, M. & Smith, P. R. *J. molec. Biol.* **129**, 359-373 (1979).
- Hendrix, R. W. *J. molec. Biol.* **129**, 375-392 (1979).
- Politt, S. & Zalkin, H. *J. Bact.* **153**, 27-32 (1983).
- Wolfe, P. B. & Wickner, W. *Cell* **36**, 1067-1072 (1984).
- Randall, L. L. & Hardy, S. J. S. *Cell* **46**, 921-928 (1986).
- Eilers, M. & Schatz, G. *Nature* **322**, 228-232 (1986).
- Crooke, E. & Wickner, W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5216-5220 (1987).
- Randall, L. L., Hardy, S. J. S. & Thom, J. R. A. *Rev. Microbiol.* **41**, 507-541 (1987).
- Griko, Yu. V., Privalov, P. L., Venyaminov, S. Yu. & Kutysenko, V. P. *J. molec. Biol.* **202**, 127-138 (1988).
- Collier, D. N., Bankaitis, V. A., Weiss, J. B. & Bassford, P. J. *Cell* **53**, 273-283 (1988).
- Neidhardt, F. C., Van Bogelen, R. A. & Vaughn, V. A. *Rev. Genet.* **18**, 295-329 (1984).
- Kochan, J. & Murialdo, H. *Virology* **131**, 100-115 (1983).
- Muller, M. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7421-7425 (1984).
- Pelman, H. R. B. & Jackson, R. J. *Eur. J. Biochem.* **67**, 247-256 (1976).

Hormone binding globulins undergo serpin conformational change in inflammation

P. A. Pemberton, P. E. Stein, M. B. Pepys*, J. M. Potter† & R. W. Carrell

Department of Haematology, University of Cambridge, Hills Road, Cambridge CB2 2QL, UK

* Department of Medicine, Royal Postgraduate Medical School, Du Cane Road, London W12 0HS, UK

† Department of Pharmacology, University of Western Australia, Nedlands, WA 6000, Australia

A surprising recent finding is that thyroxine binding globulin (TBG)¹ and cortisol binding globulin (CBG)², are members of the serine protease inhibitor (serpin) superfamily³. Apparently evolution has completely adapted the serpin structure for its new role in these proteins as a transport agent, as there is no evidence of any retained protease inhibitory activity. This drastic change in function raises the question as to why such a complex molecular framework has been selected for the relatively simple task of hormone transport? To function as inhibitors the serpins have a native stressed (S) conformation that makes them vulnerable to proteolytic cleavage, the cleavage being accompanied by an irreversible transition to a stable relaxed (R) form^{4,5}. We demonstrate here that TBG and CBG have retained the stressed native structure typical of the inhibitor members of the family and we provide evidence that the S-R transition has been adapted to allow altered hormone delivery at inflammatory sites.

TBG and CBG each have a greater than 30% homology with the archetype of the serpin family α_1 -antitrypsin, and sequence matching infers that they all share a common secondary and tertiary structure⁶. This matching is based on the crystallographic structure of α_1 -antitrypsin which has been cleaved at its reactive centre⁴. A feature of this structure is the separation of the two amino-acid residues at the reactive centre to opposite poles of

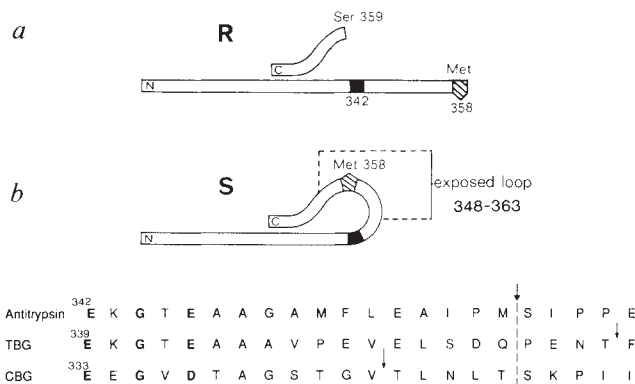


Fig. 1 The structure of the serpins is based on the cleaved (R) form of α_1 -antitrypsin shown diagrammatically (a). The native stressed (S) form has to be reconstructed by a deduced exposed loop that hinges near residue 342 (b). Alignment of the sequences of the exposed loop (based on hinge residues, bold) shows, arrowed, the sites of cleavage of TBG and CBG by neutrophil elastase. The dashed vertical line shows the putative reactive centres of TBG and CBG based on alignment with that of antitrypsin (arrowed). Human TBG and CBG were isolated from ascitic fluid and plasma respectively by affinity chromatography. Incubation with human neutrophil elastase was at 37 °C, pH 7.4 with a TBG:elastase ratio of 200:1, CBG:elastase ratio of 10:1, for 5 min. C-terminal 4K peptides were isolated by reverse-phase HPLC with a 5-80% acetonitrile gradient with 0.1% trifluoroacetic acid.

Sequencing was by the automated gas-phase method.

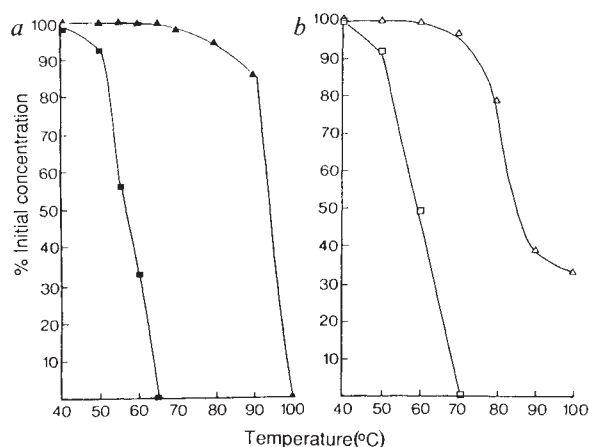


Fig. 2 Thermal stability of *a*, TBG (■, native TBG; ▲, cleaved TBG) and *b*, CBG (□, native CBG; △, cleaved CBG) showing the marked increase in stability that follows cleavage. Solutions of each protein (3 μ M in 75-mM Tris-glycine buffer, pH 7.5) in screw-capped tubes were heated at constant temperatures between 40 and 100 °C for 2 h (ref. 5). After centrifugation residual protein concentration was measured by electroimmunoassay.

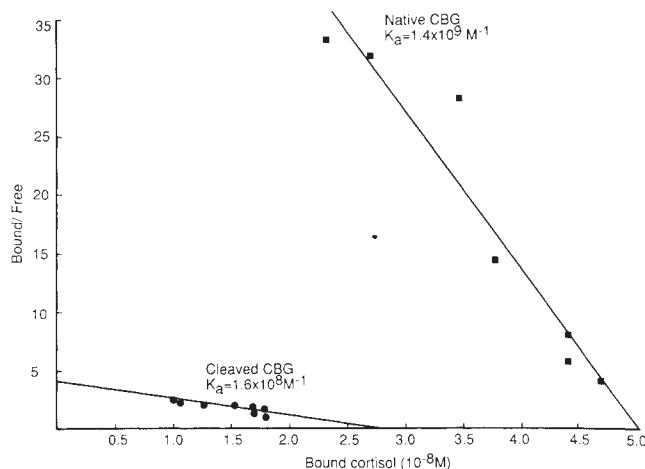


Fig. 3 Scatchard plot⁹ for CBG showing the decrease in cortisol binding affinity subsequent to cleavage. Equilibrium dialysis was performed at 4 °C in phosphate buffer pH 7.4. Dialysis bags contained 3 ml CBG (70 nM) and the outer phase (10 ml) contained [1,2-³H] cortisol (20 nCi ml⁻¹) and unlabelled cortisol (1–22 nM). After equilibration with shaking for 72 h, aliquots of inside and outside phases were counted for radioactivity. Results were corrected for non-specific binding in the presence of a 100-fold molar excess of cortisol over CBG and for the presence of endogenous cortisol in the CBG preparation.

the molecule at a distance of 69 Å. To remodel the reactive centre, as it is presumed to be in the native molecule, it is necessary to extract a peptide strand (A4) from the major (A) pleated sheet of the molecule to form an externally situated peptide loop (Fig. 1). Experimental evidence based on the inhibitor members of the family, consistently supports this model^{6,7}. The native serpins are readily cleaved, by a range of proteases, in the peptide segment corresponding to strand A4. Furthermore, the cleavage of this exposed segment is accompanied by a profound change in thermal stability, as predicted from the model. The native serpins are all relatively heat-labile, precipitating on incubation at 60 °C or less. After cleavage they remain stable in solution at temperatures of 80 °C or more³. This proteolytically induced S–R transition therefore provides a good confirmatory test of serpin tertiary structure.

We have applied this test to TBG and CBG using human neutrophil elastase to give loop cleavage. This protease is released in high concentrations by neutrophil leukocytes at sites of inflammation and it thus provides a relevant physiological model. Both TBG and CBG were incubated with neutrophil elastase and electrophoresis showed in each case that complete cleavage had occurred within 5 min, with the release of a peptide (relative molecular mass (M_r) of 4,000 (4K)), and an accompanying change in mobility of the main band, typical of that observed on cleavage of the reactive loop of serpins^{7,8}. This was confirmed by amino-acid sequencing of the released 4K polypeptides which showed cleavage within the exposed loop as in Fig. 1.

The heat stability of the cleaved TBG and CBG were then each determined over a range of temperatures and compared

with that of the native (uncleaved) forms as shown in Fig. 2. The marked increase in stability, together with the sites of cleavage, provide clear evidence that TBG and CBG do have the tertiary structure typical of serpins, with an exposed 'reactive centre' loop and a transition from S–R stability on cleavage of that loop.

What then are the advantages to these hormone carriers in retaining the S–R transition? Does cleavage result in a lower binding affinity and hence release of the hormone at inflammatory sites? To assess this, the association constants of the native and cleaved forms of TBG (for thyroxine) and CBG (for cortisol) were determined by equilibrium dialysis and Scatchard analysis⁹. There was no detectable difference in the affinity of the two forms of TBG for thyroxine. However, CBG showed a tenfold decrease in affinity for cortisol subsequent to cleavage by neutrophil elastase (Fig. 3).

On the basis of these results we can propose reasons why evolution has chosen the serpin molecule for adaptation as a hormone carrier. Its apparent complexity may be advantageous in that the S–R transition provides a switch which can be used to modulate hormone delivery. This change in conformation will result in a direct uptake of the cleaved carriers by cellular receptors which are known to be specific for the cleaved serpin conformation¹⁰. In the case of CBG, diffusion through an inflammatory locus will result in cleavage and direct release of cortisol with the potential advantage of modifying the inflammatory response and contributing to its localization¹¹.

This work was supported by the Medical Research Council and the Tobacco Advisory Council. We thank Dr L Packman for the amino-acid sequencing.

Received 2 September; accepted 13 October 1988.

1. Flink, I. L., Bailey, T. J., Gustafson, T. A., Markham, B. E. & Morkin, E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7708–7712 (1986).
2. Hammond, G. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5153–5157 (1987).
3. Carrell, R. W. & Boswell, D. R. in *Proteinase Inhibitors* (eds Barrett, A. J. & Salvesen, G.) 403–420 (Elsevier, Amsterdam, 1986).
4. Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. *J. molec. Biol.* **177**, 531–556 (1984).

5. Carrell, R. W. & Owen, M. C. *Nature* **317**, 730–732 (1985).
6. Carrell, R. W., Pemberton, P. A. & Boswell, D. R. *Cold Spring Harb. Symp. quant. Biol.* **52**, 527–535 (1987).
7. Kress, L. F. *J. cell. Biochem.* **32**, 51–58 (1986).
8. Salvesen, G. S., Catanesi, J. J., Kress, L. F. & Travis, J. J. *biol. Chem.* **260**, 2432–2436 (1985).
9. Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660–672 (1949).
10. Fuchs, H. E., Michalopoulos, G. K. & Pizzo, S. V. *J. cell. Biochem.* **25**, 231–243 (1984).
11. Parrillo, J. E. & Fauci, A. S. *A. Rev. Pharmac. Tox.* **19**, 179–201 (1979).

Oncogenic potential of *bcl-2* demonstrated by gene transfer

John C. Reed, Michael Cuddy, Trina Slabick,
Carlo M. Croce* & Peter C. Nowell

Department of Pathology and Laboratory Medicine,
University of Pennsylvania School of Medicine, Philadelphia,
Pennsylvania 19104-6082, USA

*The Wistar Institute, Philadelphia, Pennsylvania 19104-4268, USA

Follicular lymphoma is the most common human B-cell malignancy in the United States and Western Europe¹. Most of the tumours contain t(14;18) chromosome translocations involving the human *bcl-2* gene²⁻⁵. Translocation of *bcl-2* sequences from chromosome 18 into the transcriptionally active immunoglobulin locus at chromosome band 14q32 in B cells deregulates *bcl-2* gene expression, resulting in the accumulation of high levels of *bcl-2* messenger³. Human *bcl-2* transcripts generate two proteins, p26 *bcl-2*- α and p22 *bcl-2*- β , by virtue of alternative splice-site selection². Both proteins have in common their first 196 NH₂-terminal amino acids but share little similarity with other sequences in a data bank^{2,4}. Although the biological and biochemical functions of *bcl-2* are unknown, recent subcellular localization studies indicate that p26 *bcl-2*- α associates with cellular membranes, consistent with a stretch of hydrophobic amino acids in its carboxy terminus^{1,2,6}. The *bcl-2* gene may represent a novel oncogene having no known retroviral counterpart. Here we demonstrate the oncogenic potential of *bcl-2* through a gene transfer approach.

DNA sequences encoding the p26 *bcl-2*- α or p22 *bcl-2*- β were subcloned in both orientations into the pSV2-CAT expression vector⁷ to yield plasmids, pSV2-*bcl-2*- α , pSV2-*bcl-2*- α -AS (antisense, reversed orientation), pSV2-*bcl-2*- β and pSV2-*bcl-2*- β -AS (Fig. 1). Plasmid DNAs were then cotransfected with plasmids containing a dominant selectable marker (G418-resistance gene) into NIH3T3 cells (Table 1). These immortalized fibroblasts have been used extensively in assays for oncogenes by DNA transfection⁸. In these experiments, transfection of pEJ6.6 (activated human Ha-ras gene) or pSV2-c-myc-1 (murine c-myc oncogene) served as positive controls⁹, whereas transfections with pSV2-*bcl-2*- α -AS, pSV2-*bcl-2*- β -AS and pSV2gpt provided negative controls.

Monolayers of 3T3 cells transfected with an activated *ras* oncogene contained foci of morphologically transformed cells, consistent with previous reports (Table 1). In contrast, none of the other transfected plasmids, including pSV2-*bcl-2*- α and pSV2-*bcl-2*- β , induced *in vitro* morphological transformation in this assay, despite the appearance of many G418-resistant colonies. NIH3T3 cells stably transfected with pSV2-*bcl-2* plasmids, however, did produce tumours when injected into mice, with a latency of ~4 weeks. These tumours invariably grew to large size and, when allowed to progress, were lethal within 8–12 weeks.

Table 1 also summarizes the results of molecular analysis of the induced tumours. High-molecular-weight genomic DNA was isolated directly for Southern blot analysis (see Fig. 2). As shown, 13 out of 13 (100%) tumours derived from *bcl-2*- α -transfected cells and seven out of eight (88%) tumours resulting from *bcl-2*- β -transfected cells contained integrated pSV2-*bcl-2*- α and pSV2-*bcl-2*- β plasmids, respectively (Table 1). Use of a hybridization probe specific for *bcl-2* sequences (probe B) demonstrated that all *bcl-2*- α -mediated tumours contained an intact *Hind*III fragment of ~0.9 kilobase pairs (kbp) excluding the possibility that *in vivo* rearrangement of plasmid constructs accounted for the tumorigenicity (Fig. 2b). Likewise, all tumours derived from pSV2-*bcl-2*- β -transfected cells contained the expected 7.8-kbp *Hind*III fragment.

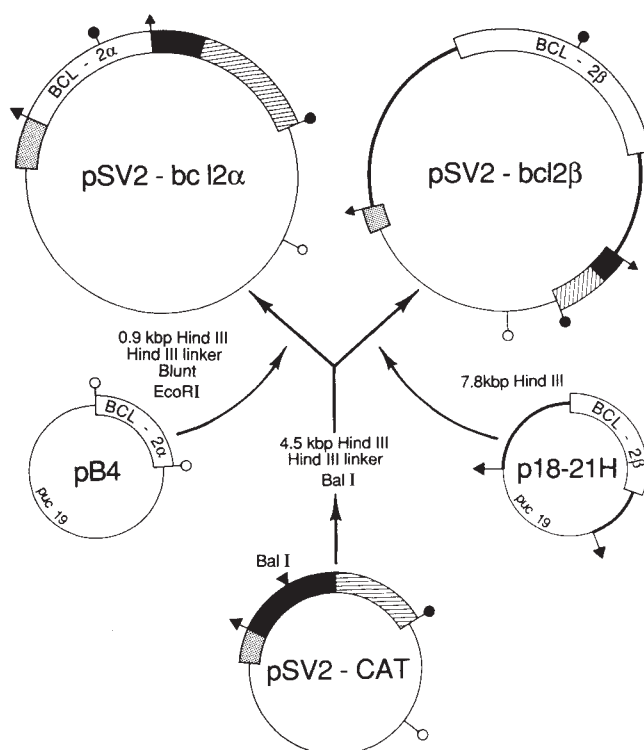


Fig. 1 Construction of plasmids expressing *bcl-2*- α and *bcl-2*- β . **Methods.** For constructions in the pSV2 expression vector, pSV2-CAT was linearized by digestion with *Bal*I, and *Hind*III-linker was added. After digestion with *Hind*III, the larger fragment (4.5 kbp) was gel-purified and ligated to either *bcl-2*- α cDNA or *bcl-2*- β genomic sequences. *Bcl-2*- α cDNA² was prepared by sequentially excising this 0.91-kbp cDNA (pB4) from its pUC19 vector with *Eco*RI, producing blunt ends with Klenow fragment, and adding *Hind*III linkers. For *bcl-2*- β , a 7.8-kbp *Hind*III-genomic clone (p18-21H) was subcloned directly into the *Hind*III site of the modified (4.5-kbp) pSV2-CAT vector. The p18-21H genomic clone contains the 3.5-kbp exon from the human *bcl-2* gene that encodes the *bcl-2*- β open reading frame (ORF)². Insert orientations were determined by extensive restriction mapping. Restriction sites for *Hind*III, *Bam*HI and *Eco*RI are denoted by $\uparrow$, $\uparrow$ and $\uparrow$, respectively. The pSV2-CAT plasmid drives expression from the SV40 early region promoter (■) and contains an intron and polyadenylation signal sequence for constitutive expression of the inserted *bcl-2* sequences (▨). A portion of the chloramphenicol acetyl transferase (CAT) gene (■) was retained in these plasmid constructions, in some cases resulting in the production of *bcl-2*-CAT fusion transcripts.

Hybridization of Southern blots with a *bcl-2*- α -CAT probe (probe A), capable of detecting restriction fragments representing junctions between integrated plasmids and adjacent cellular DNA, allowed us to estimate the number of individual clones of injected 3T3 cells present in *bcl-2*-associated tumours. As shown in Fig. 2a, all tumours contained several distinct junction fragments that were either undetectable in the polyclonal population of transfected 3T3 cells before injection or only faintly present. Assuming one integration site per stably transfected cell¹⁰, these tumours therefore developed from a minimum of four to nine individual clones (~4% of input), consistent with an oligoclonal origin. The oligoclonal nature of these tumours was subsequently confirmed by Southern blot analysis of individual clones of cells isolated from tumours whose cells were reestablished in culture (data not shown). This suggests that some minority population of transfected 3T3 cell clones acquired an additional growth advantage after injection into mice. Oligoclonal tumours were also reported recently for B lymphocytes into which both *bcl-2* and *c-myc* genes had been introduced¹¹.

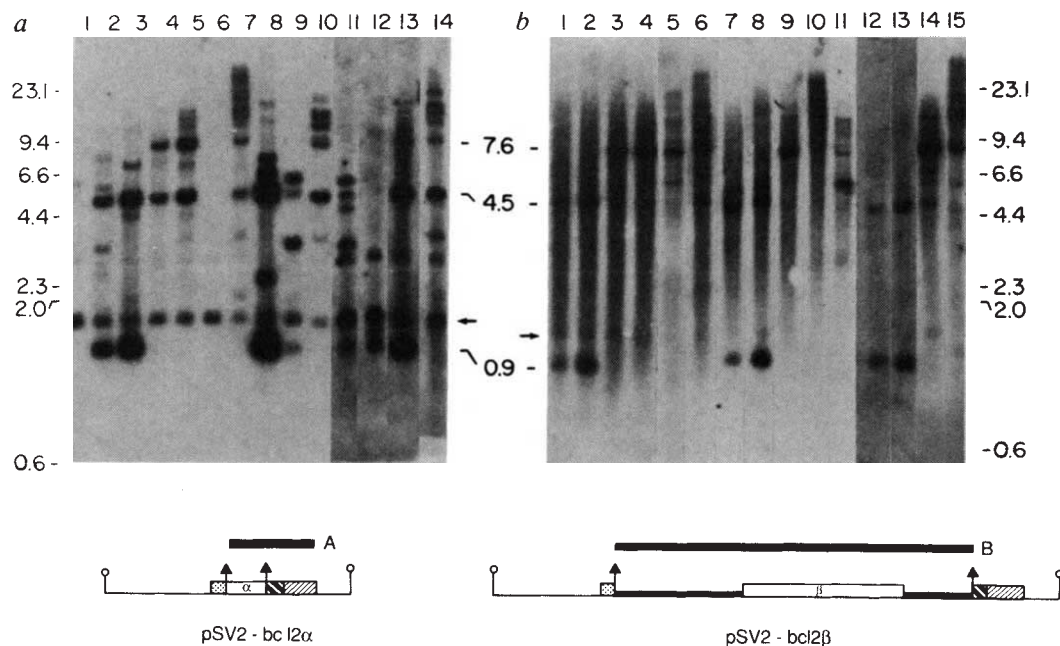


Fig. 2 Southern blot analysis of tumours.

Methods. High-molecular-weight DNA was isolated from G418-resistant 3T3 cells or from tumours, and 10 µg was digested with *Hind*III before size-fractionation in 0.8% agarose gels and transfer to nitrocellulose, essentially as described¹⁵. Blots were hybridized with ³²P-labelled probes (shown at bottom) at high stringency, washed and exposed overnight to X-ray film. Sizes of molecular-weight markers and specific restriction fragments are in kilobase pairs (kbp). The arrow indicates endogenous mouse *bcl-2* sequences. *a*, Blots were hybridized with probe A. Lanes 1–6 correspond to uninjected 3T3 cells stably transfected with pSV2gpt, pSV2-*bcl-2*-α, pSV2-*bcl-2*-α-AS, pSV2-*bcl-2*-β-AS, pSV2-*bcl-2*-β and pSVc-myc-1, respectively. Lanes 7–14 represent DNAs obtained from tumour tissues, with lanes 8, 9, 11, 12 and 13 corresponding to DNAs derived from *bcl-2*-α-mediated tumours, and lanes 7, 10 and 14 representing DNAs from *bcl-2*-β-associated tumours. *b*, Blots were hybridized with probe B. Lanes corresponding to DNAs isolated from uninjected 3T3 cells are 1, 7, 12, *bcl-2*-α (three independent transfections); 4, 9 and 14, *bcl-2*-β (three separate transfections); and lane 3, pSV2-*bcl-2*-β-AS. DNAs from *bcl-2*-α-associated tumours are in lanes 2, 8 and 13, and DNAs from *bcl-2*-β-mediated tumours are in lanes 5, 6, 10, 11 and 15. Note that lanes 1 and 2 in *b* represent 3T3 cells from the same transfection before (uninjected) and after (tumour), respectively, injection into mice. Similarly, comparisons can be made of DNAs from cellular populations before and after injection into animals for lanes 4–6, lanes 7 and 8, lanes 9–11, lanes 12 and 13, and lanes 14 and 15. Probe B contains a repetitive element accounting for high background. The 4.5-kbp band visualized in autoradiographs probably represents cross-hybridization with vector sequences (rather than contamination of probe with vector sequences), because junction fragments were not observed in DNAs from *bcl-2*-α-initiated tumours.

Fig. 3 Northern blot analysis of tumours.

Methods. Polyadenylated RNA was purified from stably transfected 3T3 cells and from portions of tumours, and 5–10 µg was size-fractionated in 6% formaldehyde–1% agarose gels before transfer to nylon membranes as described elsewhere¹⁶. RNA blots were hybridized with ³²P-probe A (top) (see Fig. 2), washed and exposed to X-ray film overnight. Sizes of mRNAs (kbp) were estimated from an RNA ladder (BRL, Bethesda, Maryland) run in gels and visualized by staining with ethidium bromide (data not shown). Shown are RNAs from *a*, 3T3 cells stably transfected with pSV2-*bcl-2*-β (lane 1) or pSV2-*bcl-2*-α (lane 2); *b*, 3T3 cells transfected with pSV2-*bcl-2*-α (lane 3), pSV2-*bcl-2*-α-AS (lane 4), or pSV2-*bcl-2*-β-AS (lane 5); *c*, *bcl-2*-α-mediated tumours (lanes 6–9) showing examples of tumours that contained both 2.4 and 1.4-kbp transcripts (lane 6), predominantly 2.4-kbp mRNA (lane 7), mostly 1.4-kbp transcripts (lane 8), and no detectable *bcl-2* transcripts (lane 9); *d*, *bcl-2*-β-associated tumours (lanes 10, 11) showing examples of tumour tissue containing no detectable *bcl-2* mRNA (lane 10) and a tumour containing a major 3.5-kbp mRNA and a minor 6.2-kbp transcript (lane 11); and *e*, tumour cells from lanes D9 (*bcl-2*-α) and E10 (*bcl-2*-β) expanded in culture to achieve a more homogeneous population of tumour cells before isolation of RNA and reanalysis by RNA blotting (see legend to Table 1). Among *bcl-2*-α-mediated tumours, five (45%) expressed predominantly a 2.4-kbp transcript, three (27%) contained predominantly a 1.4-kbp mRNA, and three expressed both the 2.4- and 1.4-kbp mRNAs. Rehybridization of blots with a murine β-2-microglobulin probe (bottom) controlled for variations in amounts of RNA samples on blots¹⁷.

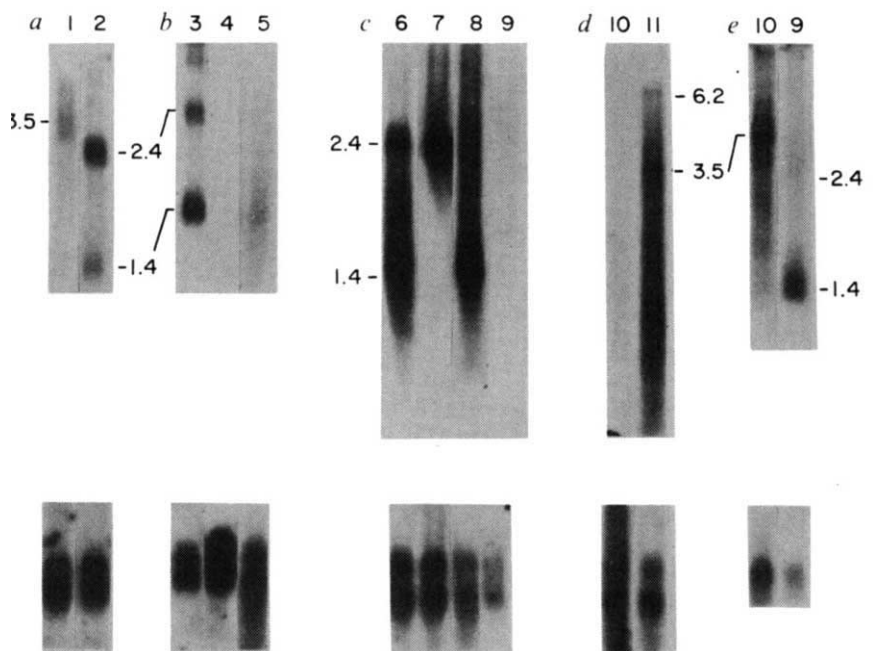


Table 1 Summary of tumorigenicity studies for transfected NIH3T3 fibroblast cells

Plasmid	Foci/neo ^R colonies	Tumours/injections (%)	Mean time weeks (range)	Tumours with DNA (%)	Tumours with RNA (%)
pSV2gpt	0/82	0/4 (0)	NA	NA	NA
pSV2- <i>bcl-2-α</i>	0/94	13/14 (93)	4.2 (3–6)	13/13 (100)	11/11 (100)*
pSV2- <i>bcl-2-α</i> -AS	1/63	1/7 (14)†	5.0 (5)	0/1 (0)	ND
pSV2- <i>bcl-2-β</i>	0/264	8/9 (89)	3.7 (3–4)	7/8 (88)	6/6 (100)*
pSV2- <i>bcl-2-β</i> -AS	0/86	0/4 (0)	NA	NA	NA
pSV-2- <i>c-myc</i> -1	1/116	2/4 (50)‡	5.5 (5–6)	2/2 (100)	ND
pEJ6.6	202/378	2/2 (100)	2.0 (2)	2/2 (100)	ND

NIH3T3 fibroblasts (10^6 cells per 100-mm dish) were cotransfected with 10–15 µg of various plasmid DNAs as shown and 2–3 µg of a plasmid containing a neomycin-resistance gene (either pCDneo or pBC140)^{13,14} using a calcium phosphate precipitation method⁷. Two days later, transfected cells were reseeded 1:20 and replicate cultures were fed twice weekly with fresh medium either containing or not containing 800 µg ml⁻¹ (active concentration ~400 µg ml⁻¹) G418 (Geneticin; Gibco/BRL). After 10–14 days, plates were examined microscopically for the presence of foci of morphologically transformed cells in monolayer cultures not subjected to antibiotic selection, and for colonies of antibiotic-resistant cells in cultures treated with G418. Antibiotic-resistant 3T3 cells were expanded in culture in the presence of G418, and 5×10^6 cells were injected subcutaneously into irradiated (300 rad) 6–8-week-old female athymic mice (Harlan-Sprague-Dawley, Indianapolis, IN). Mice were observed weekly for 8 weeks for macroscopic evidence of tumour formation. Tumour tissues were analysed by Southern and Northern blotting for presence of stably transfected plasmid DNAs and for construct expression, respectively. See text and Fig. 1 for description of plasmids. The results may be dependent on the subclone of NIH3T3 cells used, because in similar experiments (data not shown) one of three sources of 3T3 cells tested failed to yield tumours with *bcl-2*.

* Northern blot analysis of RNAs isolated directly from excised tumours revealed that 10 out of 19 *bcl-2*-induced tumours (seven out of 11 *bcl-2-α* and three out of eight *bcl-2-β*) contained detectable levels of recombinant *bcl-2* transcripts. Because excised tumours contain non-neoplastic cells whose RNA may dilute the RNA from tumour cells, we propagated in culture cells from seven out of the nine tumours that lacked evidence for expression of *bcl-2*, hoping to minimize the non-neoplastic component. RNA isolated from all cultured tumours (four *bcl-2-α*, three *bcl-2-β*) contained detectable recombinant *bcl-2* transcripts (see lanes 9 and 10 of Fig. 3, for example). Intact RNA was not obtained from two of the 13 *bcl-2-α*-mediated tumours. Two of the eight *bcl-2-β*-mediated tumours could not be propagated *in vitro* for further RNA blot analyses (see text).

† 3T3 cells transfected with pSV2-*bcl-2-α*-AS formed a small tumour in one case that never grew >0.5 cm in diameter.

‡ The *c-myc* construct pSV-*c-myc*-1 was weakly tumorigenic in our experiments, though it complemented efficiently with an activated *ras* gene to transform rat embryo fibroblasts (data not shown)⁹. RNA blot analysis of 3T3 cells transfected with pSV-*c-myc*-1 revealed a low level of *c-myc* expression from this construct. As reported previously^{8,9,12}, *c-myc* did not induce morphological transformation of 3T3 cells. NA = not applicable; ND = not done.

Northern blot analysis of RNAs isolated from tumour cells revealed that, in all cases where a complete analysis was possible (11 for *bcl-2-α*, six for *bcl-2-β*), these tumours contained detectable levels of recombinant *bcl-2* transcripts (Table 1). Furthermore, relative levels of recombinant *bcl-2* transcripts generally appeared higher in tumour-derived cells that were reestablished in culture (Fig. 3c–e) than in the original 3T3-transfectants (Fig. 3a, b), suggesting an *in vivo* growth advantage for cells expressing higher levels of *bcl-2*. We cannot determine from this analysis whether *in vivo* selection for high levels of *bcl-2* expression contributed to the oligoclonal origin of these tumours (Fig. 2).

Cells transfected with the pSV2-*bcl-2-β* plasmid contained a predominant mRNA of 3.5 kbp corresponding to the normal human *bcl-2-β* transcript² (Fig. 3a, d, e). Occasionally, low levels of a 6.2 kbp transcript consistent with initiation from the SV40 promoter were also detected (see lane 11 of Fig. 3e for example). Cells stably transfected with pSV2-*bcl-2-α* expressed two transcripts of ~1.4 kbp and 2.4 kbp (see Fig. 3a, b). Preliminary mapping of these transcripts by RNA protection assay indicates that the 2.4-kbp mRNA contains intact the *bcl-2-α* coding region, whereas the 1.4-kbp transcript may lack some 5' *bcl-2-α* complementary DNA sequences (data not shown). We detected only the full-length 2.4-kbp transcripts in five (45%) out of the 11 tumours generated from 3T3 cells transfected with pSV2-*bcl-2-α* and analysed by RNA blotting, indicating that production of the truncated transcript is not necessary for *bcl-2-α*-mediated tumorigenicity (see lane 7 of Fig. 3c for example).

Though the *bcl-2* gene has been strongly implicated in the aetiology of human follicular lymphoma^{2–5}, formal evidence for its oncogenic potential has been lacking. Here we have demonstrated that *bcl-2* sequences can contribute to the tumorigenicity of NIH3T3 cells. Transfection of *bcl-2* constructs specifically enhanced tumorigenicity in these cells, because mice injected with NIH3T3 cells transfected with negative control plasmids

failed to develop tumours in our experiments. The latency period required for *bcl-2*-associated tumour formation (~four weeks) is similar to that reported for the p53 tumour antigen¹², providing further evidence for an important role for *bcl-2* production in the observed tumour development.

Unlike many other oncogenes that encode membrane-associated proteins (for example, *ras* and tyrosine kinase-encoding oncogenes such as *erbB*)⁸, *bcl-2* did not directly induce transformation of 3T3 cells *in vitro* (Table 1). But cells recovered from *bcl-2*-mediated tumours exhibited a transformed morphology when reestablished in culture (data not shown), indicating that additional secondary events probably contributed to their tumorigenicity. Thus, our *bcl-2* constructs either provided only some of the molecular changes necessary for tumour formation, or indirectly enhanced tumorigenicity of 3T3 fibroblasts. Regardless of its mechanism of action, *bcl-2* is the first example of a novel oncogene to be identified by virtue of its involvement in chromosomal translocations in human tumours.

Received 16 August; accepted 11 October 1988.

1. Jones, S. E. *et al.* *Cancer* **31**, 806–823 (1973).
2. Tsujimoto, Y. & Croce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5214–5218 (1986).
3. Tsujimoto, Y., Cossman, J., Jaffe, E. & Croce, C. M. *Science* **228**, 1440–1443 (1985).
4. Cleary, M. L., Smith, S. D. & Sklar, J. *Cell* **47**, 19–28 (1986).
5. Seto, M. *et al.* *EMBO J* **7**, 123–131 (1988).
6. Tsujimoto, Y., Ikegaki, N. & Croce, C. M. *Oncogene* **2**, (1987).
7. Gorman, C. in *DNA Cloning* Vol. II (ed. Glover, D. M.) 143–190 (IRL, Oxford, 1985).
8. Land, H., Parada, L. F. & Weinberg, R. A. *Science* **222**, 771–778 (1983).
9. Land, H., Parada, L. F., & Weinberg, R. A. *Nature* **304**, 596–602 (1983).
10. Scangos, G. & Ruddle, F. H. *Gene* **14**, 1–10 (1981).
11. Vaux, D., Cory, S. & Adams, J. *Nature* **335**, 440–442 (1988).
12. Kelekar, A. & Cole, M. D. *Molec. Cell Biol.* **6**, 7–14 (1986).
13. Chen, C. & Okayama, H. *Molec. Cell Biol.* **7**, 2745–2752 (1987).
14. Wano, Y., Cullen, B., Svetlik, P., Pfeffer, N. & Greene, W. *Molec. Biol. Med.* **4**, 1–15 (1987).
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning, A Laboratory Manual*, 383–389 (Cold Spring Harbor Press).
16. Reed, J. C., Tsujimoto, Y., Alpers, J. D., Croce, C. M. & Nowell, P. C. *Science* **236**, 1295–1299 (1987).
17. Parnes, J. & Seidman, J. *Cell* **29**, 661–669 (1982).

Retinoic acid and thyroid hormone induce gene expression through a common responsive element

Kazuhiko Umesono*, Vincent Giguere*†, Christopher K. Glass‡, Michael G. Rosenfeld‡§ & Ronald M. Evans*§

§ Howard Hughes Medical Institute, * The Salk Institute for Biological Studies, La Jolla, California 92037-0613, USA
‡ Eukaryotic Regulatory Biology Program, Center for Molecular Genetics, School of Medicine, University of California, San Diego, California 92093-0613, USA

Studies of steroid receptors have led to the identification of a superfamily of ligand-inducible regulatory proteins that includes receptors for thyroid hormones and retinoic acid¹. This family of receptors regulates gene expression through binding to short cis-acting sequences referred to as hormone-response elements. Identification of a functional retinoic acid responsive element is crucial to our understanding of the mechanisms by which retinoic acid receptors activate gene expression and regulate cell differentiation. One impediment to such a study is the absence of any identified gene whose transcription is directly dependent on the receptor-hormone complex. Because the DNA-binding domains of the retinoic acid and thyroid hormone receptors are highly related (62% identical in their amino acid sequences)², we have investigated the possibility that the retinoic acid receptor could activate gene expression through a thyroid hormone response element. We now report that a human retinoic acid receptor expressed from cloned complementary DNA or the endogenous retinoic acid receptor present in F9 teratocarcinoma cells can activate gene expression from promoters fused to a natural or synthetic thyroid hormone response element. The product translated *in vitro* from the human retinoic acid receptor cDNA can bind to a thyroid hormone response element with high affinity. The unexpected implication of these findings is that retinoic acid and thyroid hormones, acting through their respective receptors, could control overlapping gene networks involved in the regulation of vertebrate morphogenesis and homeostasis.

Novel thyroid hormone (3,5,3'-triiodo-L-thyronine, T₃) responsive promoters were constructed by replacing the glucocorticoid responsive elements present in the mouse mammary tumour virus-long terminal repeat (MTV-LTR) with an oligonucleotide encoding either a palindromic thyroid hormone response element (TRE) (TRE_p, 5'-TCAGGTCATGACCTGA 3') (ref. 3) or a natural rat growth hormone TRE (TRE_{GH}, 5'-CAGGACGTGACCGCA 3') (ref. 4). These promoters were then fused to the chloramphenicol acetyl transferase (CAT) gene to generate reporter plasmids ΔMTV-TRE_p-CAT and ΔMTV-TRE_{GH}-CAT (Fig. 1a; C. C. Thompson, K. Damm and R.M.E., unpublished results). When receptorless CV-1 cells are cotransfected with the expression vector containing a human thyroid hormone receptor⁵ (pRShT₃Rβ) and the reporter plasmid ΔMTV-TRE_p-CAT, a 40-fold induction in CAT activity is observed in the presence of T₃ (Fig. 1b). In contrast, cotransfection of an expression vector encoding the human glucocorticoid receptor (pRShGR) (ref. 6) and the same reporter plasmid does not stimulate CAT activity in response to the synthetic glucocorticoid dexamethasone. Remarkably, retinoic acid provoked a dramatic increase (20-fold) in CAT activity when the expression vector encoding the human retinoic acid receptor (hRARα) (pRShRARα) (ref. 2) was cotransfected with the TRE reporter. The results shown in Fig. 1c demonstrate that the induction of CAT activity by the hRARα is dependent on the palindromic

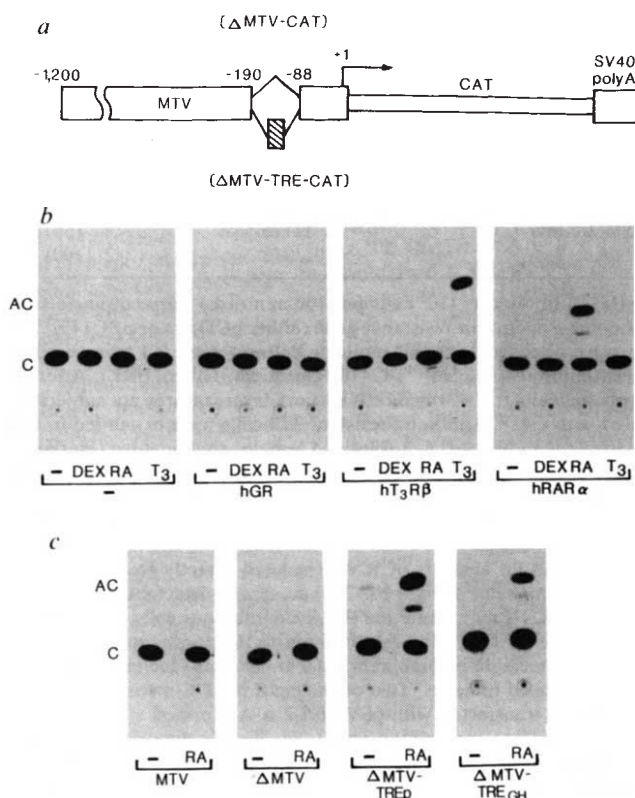


Fig. 1 a, The MTV-CAT reporter gene and its derivatives. The glucocorticoid-responsive reporter plasmid MTV-CAT contains the MTV-LTR linked to the CAT coding sequence. Numbering is relative to the start site of transcription. Major glucocorticoid response elements located in positions between -190 and -88 were deleted to make ΔMTV-CAT. To construct thyroid hormone-responsive reporter plasmids, synthetic oligonucleotides containing either a palindromic thyroid hormone response element (TRE_p) or the rat growth hormone gene TRE (TRE_{GH}) were inserted into the ΔMTV-CAT promoter at the -190/-88 fusion point to give ΔMTV-TRE_p-CAT and ΔMTV-TRE_{GH}-CAT reporter plasmids, respectively. b, Human thyroid hormone and retinoic acid receptors induce CAT activity through a common response element. Each receptor expression plasmid is cotransfected into CV-1 cells with the thyroid hormone-responsive reporter ΔMTV-TRE_p-CAT and the cells were cultured in the absence (-) or presence of 100 nM each of dexamethasone (DEX), retinoic acid (RA), or 3,5,3'-triiodo-L-thyronine (T₃). The receptors inserted into the expression vector are: hGR, human glucocorticoid receptor; hT₃Rβ, human thyroid hormone receptor-β; hRAR, human retinoic acid receptor-α; the minus signifies hT₃Rβ in the reverse orientation. C, chloramphenicol; AC, 3-acetylchloramphenicol. c, Induction of CAT activity by hRARα. Each reporter plasmid shown in a was cotransfected into CV-1 cells together with a hRARα expression plasmid and cultured in the absence (-) or presence of 100 nM retinoic acid.

Methods. a, Construction of MTV-CAT and ΔMTV-CAT has been described⁶. Construction of ΔMTV-TRE_p-CAT and ΔMTV-TRE_{GH}-CAT will be described elsewhere (C. C. Thompson, K. Damm and R.M.E., unpublished results). b and c, Cell transfection and CAT assay. The expression plasmid (1 μg), the reporter plasmid (5 μg), the reference plasmid pRSV-BGAL (5 μg), and the carrier plasmid pUC19 (9 μg) were introduced into CV-1 cells (5 × 10⁵ per 10 cm dish) by calcium phosphate coprecipitation⁹. The cells were cultured for 36 h in Dulbecco-modified Eagle medium containing 10% resin-charcoal stripped calf bovine serum with or without inducers, and subsequently prepared for CAT assays through three cycles of freeze-thawing¹⁰. Aliquots of the cell extracts were normalized by β-galactosidase activity prior to the CAT assay¹¹.

† Present address: Division of Endocrinology, Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto M5G 1X8, Canada.

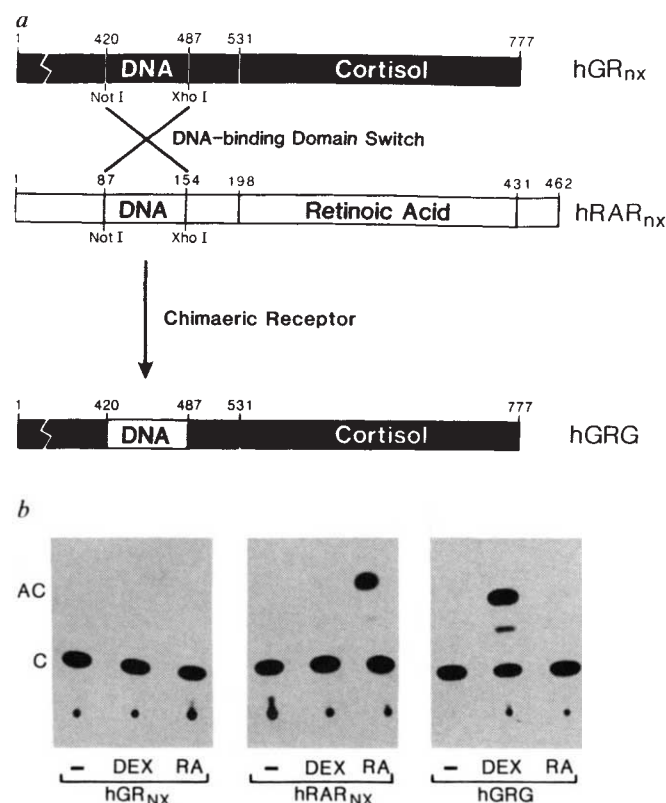


Fig. 2 *a*, Construction of the chimaeric receptor hGRG. The hGR_{NX} and hRAR_{NX} are mutated hGR and hRAR α , respectively, with common *NotI* and *XhoI* sites in the cDNAs. The amino acid numbers represent the possible domain boundaries in the receptor proteins. The ligand-binding domains are indicated by their respective inducers. The DNA-binding domains represented by 'DNA' were exchanged between hGR_{NX} and hRAR_{NX} to create the chimaeric receptor hGRG, a mutant hGR with the hRAR α DNA-binding specificity. *b*, Trans-activation of a T₃ responsive reporter by the hybrid receptor. Expression plasmids encoding mutant receptors were cotransfected into CV-1 cells together with the Δ MTV-TRE_p-CAT and cells were cultured for 36 h in the absence (-) or presence of 100 nM inducers. No effect on CAT activity was observed using the parent vector pRShGR_{NX}. When pRShRAR_{NX} was cotransfected, retinoic acid mediated a 10-fold induction. AC and C as described for Fig. 1.

Methods. Construction of hGR_{NX} and hRAR_{NX} has been described². To construct the expression plasmid RShGRG, the small *NotI*-*XhoI* fragment of pRShRAR_{NX} was isolated and ligated with the large *NotI*-*XhoI* fragment of pRShGR_{NX}. Cell transfection and CAT assay are described in the legend to Fig. 1.

or natural TREs. As previously observed with thyroid hormone receptor³, the palindromic variant TRE conferred a greater increase in gene expression than the natural TRE_{GH}.

To demonstrate that the DNA-binding domain of the hRAR α is responsible for the specific recognition of the TRE_p, we have substituted the DNA-binding domain of the hGR with that from the hRAR α to create the hGRG chimaera (Fig. 2*a*). In principle, this hybrid should confer cortisol responsiveness on genes normally activated by retinoic acid. When CV-1 cells are cotransfected with the expression plasmid encoding hGRG and the reporter Δ MTV-TRE_p-CAT, dexamethasone specifically elicited a 50-fold stimulation in CAT activity (Fig. 2*b*). This experiment provides direct evidence that the DNA-binding domain of the hRAR α determines the specificity of target gene activation.

The DNA binding specificity of the retinoic acid receptor (RAR) was examined using an avidin-biotin complex DNA-binding assay⁴. The results of Fig. 3*a* show that the hRAR α

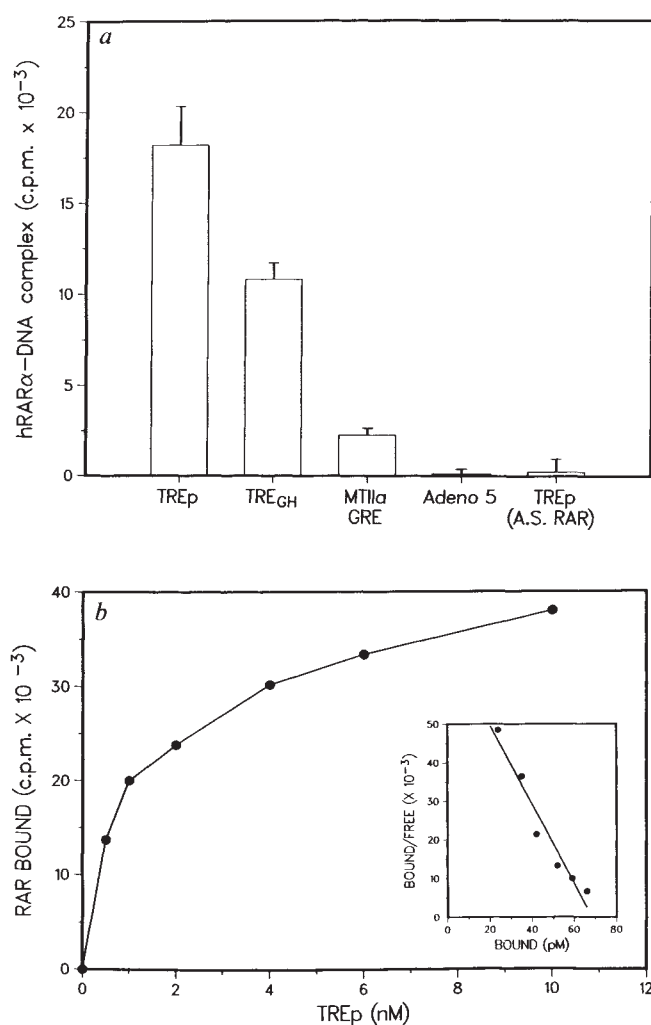
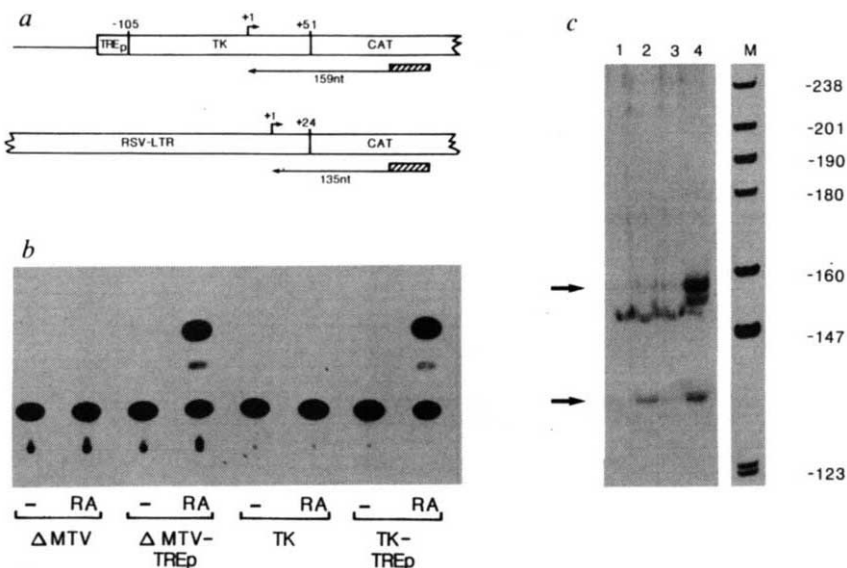


Fig. 3 DNA-binding properties of the retinoic acid receptor. *a*, Sequence-specific precipitation of [³⁵S]methionine-labelled retinoic acid receptor by oligonucleotide probes containing biotin-11-dUTP. TRE_p contains the palindromic TRE (ref. 3); TRE_{GH}, the TRE from the rat growth hormone gene⁴; MTIIaGRE, the glucocorticoid response element from the metallothionein IIa gene¹². Adeno 5 contains a *cis*-active element from adenovirus 5 (ref. 13). Radiolabelled retinoic acid receptor was prepared using capped mRNA transcripts to program translation in a rabbit reticulocyte lysate in the presence of [³⁵S]methionine. Precipitations were performed using 200 fmol of each biotinylated probe and 5 μ l reticulocyte lysate. The bar labelled A.S. RAR represents DNA-dependent radioactivity observed using 5 μ l reticulocyte lysate programmed with antisense retinoic acid receptor mRNA. Background radioactivity, representing precipitated ³⁵S counts in the absence of biotinylated oligonucleotide, was 4,300 c.p.m. Results are plotted as the mean and standard error of triplicate points and represent four experiments on the sequence specificity of retinoic acid receptor binding. *b*, Concentration dependence of hRAR α binding to the palindromic TRE. Aliquots (10 μ l) of hRAR-reticulocyte lysate were incubated with the indicated concentrations of the biotinylated palindromic TRE and precipitated with streptavidin agarose. The specific activity of [³⁵S]methionine was used to estimate the concentration of protein-DNA complexes. The inset shows a Scatchard analysis of these data.

Methods. Biotinylated oligonucleotide probes were prepared as described⁴. The indicated volumes of hRAR α reticulocyte lysate were incubated with 25 μ l buffer H (20 mM HEPES, pH 7.8, 50 mM KCl, 1 mM mercaptoethanol, 0.1% Nonidet P-40), 5 μ l poly(dI-dC) (1 mg ml⁻¹), 10 μ l F9 whole-cell extract and 10 μ l biotinylated oligonucleotide at room temperature for 30 min. Protein-DNA complexes were precipitated by addition of 20 μ l streptavidin agarose and washed 3 times with 1 ml ice-cold buffer H.

Fig. 4 Transcriptional activation of T_3 responsive reporter genes in F9 teratocarcinoma cells in response to retinoic acid. *a*, Organization of T_3 responsive and control reporter plasmids. The oligonucleotide encoding TRE_p was inserted upstream of the TK promoter present in the plasmid pBLCAT2 (ref. 14), and the new T_3 responsive reporter plasmid was designated TK-TRE_p-CAT (K. Damm, C. C. Thompson and R.M.E., manuscript in preparation). To analyse the start site of CAT mRNAs in retinoic acid-induced cells, the Rous sarcoma virus-derived RSV-CAT (ref. 15) was used as an internal control. For the primer extension analysis, a 33-mer oligonucleotide complementary to the CAT sequence (hatched boxes) was used as a primer. The extended products are 159 nucleotides long for TK-CAT and TK-TRE_p-CAT and 135 nucleotides long for RSV-CAT when transcriptional start sites are correct. *b*, Induction of CAT activity by retinoic acid in F9 teratocarcinoma cells. F9 cells were transfected with either non-responsive (Δ MTV-CAT or TK-CAT) or T_3 responsive (Δ MTV-TRE_p-CAT or TK-TRE_p-CAT) reporter plasmids, and cultured for 36 h in the absence (-) or presence of 100 nM retinoic acid. Note that no receptor expression plasmid was cotransfected. *c*, Primer extension analysis of the retinoic acid-induced CAT mRNA in F9 cells. Together with the control RSV-CAT, either TK-CAT (lanes 1 and 2) or TK-TRE_p-CAT (lanes 3 and 4) plasmids were cotransfected into undifferentiated F9 cells. The cells were then cultured for 24 h in the absence (lanes 1 and 3) or presence (lanes 2 and 4) of retinoic acid (1 μ M). Cytoplasmic RNA was then isolated and the 5' end of the CAT mRNA was analysed by the primer-extension method using the oligonucleotide described in (*a*) as a primer. Lane M shows size markers in nucleotides, and arrows indicate the primer extension products of the plasmids containing the TK promoter (159 nucleotides) and RSV-LTR (135 nucleotides). A slight induction of RSV-CAT was observed in response to retinoic acid, although this induction was not observed in CV-1 cells (V.G., unpublished observation). **Methods.** *a*, To construct TK-TRE_p-CAT plasmid, annealed TRE_p oligonucleotides were inserted at the *Hind*III site of pBLCAT2. The plasmid pRSVCAT has been described¹⁵. *b*, The reporter plasmid (10 μ g) and the carrier pGEM3 (5 μ g) were cotransfected into undifferentiated F9 cells by modified calcium phosphate coprecipitation¹⁶ and the cells were maintained in the media described in Fig. 1. Cell extracts were assayed for CAT activity using 350 μ g protein for 6 h (Δ MTV and Δ MTV-TRE_p) or 40 μ g protein for 30 min (TK and TK-TRE_p). *c*, F9 cells were cotransfected with either reporter plasmid TK-CAT or TK-TRE_p-CAT (5 μ g), the internal control plasmid RSV-CAT (5 μ g) and the carrier plasmid pGEM3 (10 μ g). Primer extension was by the method of Elsholtz *et al.*¹⁷.



efficiently binds only to biotinylated oligonucleotides containing the TRE_p or the TRE_{GH}. Consistent with the functional data demonstrating a higher level of retinoic acid responsiveness from the TRE_p, this oligonucleotide also complexes the hRAR α more efficiently than TRE_{GH}. Biotinylated oligonucleotides containing unrelated sequences, such as a *cis*-active element from the adenovirus 5 LTR, show no specific binding. A partially related sequence, the metallothionein IIa glucocorticoid responsive element, gave low but detectable binding to the hRAR α . Scatchard analysis of the concentration dependence of RAR binding to the TRE_p indicated a binding constant of ~ 1 nM (Fig. 3b), which is similar to the value for the interaction between the *in vitro*-translated T_3 R and this element (unpublished result).

At present there is no direct assay to evaluate the function of endogenous RARs. The results shown in Figs 1–3 suggest that TRE_p-containing vectors could provide a simple and quantitative assay for retinoic acid responsiveness in cultured cells. To explore this possibility, we introduced two different sets of reporters into undifferentiated F9 cells which undergo cell differentiation in the presence of retinoic acid⁷. The first is the Δ MTV-TRE_p-CAT already described (Fig. 1a), and the second contains TRE_p linked to the TK promoter (Fig. 4a; K. Damm, C. C. Thompson and R.M.E., manuscript in preparation). As shown in Fig. 4b, F9 cells transfected with either reporter plasmid containing TRE_p can induce CAT activity when exposed to retinoic acid. This induction of CAT activity is clearly mediated through the TRE_p because deletion of this sequence renders these promoters non-responsive to retinoic acid. The induction is not conferred by thyroid hormone receptors because addition of T_3 (100 nM) does not stimulate CAT activity (data not shown). CAT messenger RNA was analysed by a primer-extension technique to determine whether the retinoic acid-dependent stimulation of CAT activity resulted from an increase in appropriately

initiated transcripts. The TRE_p fused to the TK promoter produced a large increase in correctly initiated transcripts in response to retinoic acid (Fig. 4c, lane 4), a result that is in agreement with the observed induction in CAT activity. These results provide a potential strategy for the characterization of the functional properties of endogenous RAR in a variety of cultured cells.

Although retinoic acid and thyroid hormones are neither biosynthetically nor structurally related and control diverse biological events, our findings demonstrate that these hormones act through their respective receptors to activate gene expression through a common response element, and thus may control overlapping gene networks.

We thank Jeff Arriza for the plasmid RShGR_{nx}, Catherine C. Thompson and Dr Klaus Damm for the reporter plasmids TK-TRE-CAT and Δ MTV-TRE-CAT, Dr G. Schuetz for pBLCAT2, Callie Mack for oligonucleotide synthesis and Dr Yasuaki Shirayoshi for advice on F9 cell culture. V.G. is a Centennial Fellow of the Medical Research Council of Canada. C.K.G. holds a Lucille Markey Scholarship for Biomedical Science. M.G.R. and R.M.E. are investigators of the Howard Hughes Medical Institute. This work was supported by the Howard Hughes Medical Institute and by grants from the National Institute of Health.

Received 15 August; accepted 5 October 1988.

- Evans, R. M. *Science* **240**, 889–895 (1988).
- Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. *Nature* **330**, 624–629 (1987).
- Glass, C. K., Holloway, J. M., Devari, O. & Rosenfeld, M. G. *Cell* **54**, 313–323 (1988).
- Glass, C. K. *et al.* *Nature* **329**, 738–741 (1987).
- Weinberger, C. W. *et al.* *Nature* **324**, 641–646 (1986).
- Giguere, V., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. *Cell* **46**, 645–652 (1986).
- Strickland, S. & Mahdavi, V. *Cell* **15**, 393–403 (1978).
- Hollenberg, S. M. & Evans, R. M. *Cell* **55**, (in the press).
- Wigler, M. *et al.* *Cell* **16**, 777–785 (1979).

10. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
11. Herbomel, P., Bourachot, B. & Yaniv, M. *Cell* **39**, 653-662 (1984).
12. Karin, M. *et al. Nature* **308**, 513-519 (1984).
13. Hearing, P. & Shenk, T. *Cell* **33**, 695-703 (1983).
14. Luckow, B. & Schuetz, G. *Nucleic Acids Res.* **15**, 5490 (1987).
15. Gorman, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 6777-6781 (1982).
16. Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745-2752 (1987).
17. Elsholtz, H. P. *et al. Science* **234**, 1552-1557 (1986).

The role of the distal histidine in myoglobin and haemoglobin

John S. Olson*, Antony J. Mathews*,
 Ronald J. Rohlfs*, Barry A. Springer†,
 Karen D. Egeberg†, Stephen G. Sligar†,
 Jeremy Tame‡, Jean-Paul Renaud‡§
 & Kiyoshi Nagai‡

* Department of Biochemistry, Rice University, Houston, Texas 77251, USA

† Departments of Biochemistry and Chemistry, University of Illinois, Urbana, Illinois 61801, USA

‡ MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The distal E7 histidine in vertebrate myoglobins and haemoglobins has been strongly conserved during evolution and is thought to be important in fine-tuning the ligand affinities of these proteins¹⁻⁸. A hydrogen bond between the N_ε proton of the distal histidine and the second oxygen atom may stabilize O₂ bound to the haem iron¹⁻⁸. The proximity of the imidazole side chain to the sixth coordination position, which is required for efficient hydrogen bonding, has been postulated to inhibit sterically the binding of CO and alkyl isocyanides²⁻⁸. To test these ideas, engineered mutants of sperm whale myoglobin⁹ and the α - and β -subunits of human haemoglobin^{8,10-12} were prepared in which E7 histidine was replaced by glycine. Removal of the distal imidazole in myoglobin and the α -subunits of intact, R-state haemoglobin caused significant changes in the affinity for oxygen, carbon monoxide and methyl isocyanide; in contrast, the His-E7 to Gly substitution produced little or no effect on the rates and extents of O₂, CO and methyl isocyanide binding to β -chains within R-state haemoglobin. In the β -subunit the distal histidine seems to be less significant in regulating the binding of ligands to the haem iron in the high affinity quaternary conformation. Structural differences in the oxygen binding pockets shown by X-ray crystallographic studies^{4,5} account for the functional differences of these proteins.

Wild-type and mutant sperm whale myoglobins were produced in *Escherichia coli* from a synthetic gene as described previously⁹. In this expression system, the holoprotein is synthesized directly by the bacterium and can be isolated in a soluble, fully functional form. Wild-type and mutant α - and β -chains of haemoglobin were expressed independently in *E. coli*¹⁰⁻¹². In this case, the globin was produced as a fusion protein. After liberation of the N-terminal leader sequence by specific proteolysis, the globin was reconstituted with haemin dicyanide in the presence of the corresponding native subunit partner prepared in the reduced CO form from human blood. After reduction and purification, the final product was tetrameric haemoglobin containing one pair of mutant subunits and one pair of normal subunits. Comparisons of some of the kinetic properties of native, *E. coli* wild-type and *E. coli* mutant haemoglobins and myoglobins are shown in Table 1 and confirm the integrity of synthetically produced proteins.

We examined only the properties of R-state haemoglobin subunits because T-state parameters have to be obtained

Table 1 Comparison of the ligand dissociation rate constants of native, *E. coli* wild-type and *E. coli* mutant haemoglobins and myoglobins at pH 7, 20 °C

Protein	Oxygen k (s ⁻¹)	Carbon monoxide k (s ⁻¹)	Methyl isocyanide k (s ⁻¹)
α (Native His-E7)	9.3	0.0050	1.8
β (Native His-E7)	19.5	0.0100	5.6
α (<i>E. coli</i> His-E7)	11.6	0.0043	1.7
β (Native His-E7)	24.2	0.0066	7.0
α (<i>E. coli</i> Gly-E7)	623	0.0067	0.28
β (Native His-E7)	21.7	0.0067	5.8
α (Native His-E7)	10	0.0052	1.3
β (<i>E. coli</i> Gly-E7)	37	0.0130	4.2
Mb (Native His-E7)	12	0.019	4.3
Mb (<i>E. coli</i> His-E7)	20	0.018	4.3
Mb (<i>E. coli</i> Gly-E7)	1,620	0.038	6.3

Nagai and co-workers have shown previously that *E. coli*, wild-type (His-E7) β -chains can be successfully reconstituted with native α -subunits to produce haemoglobin molecules which show oxygen-binding parameters indistinguishable from native haemoglobin A (HbA)⁷. The preparation of the corresponding α (*E. coli* wild-type)₂ β (native)₂ protein followed that previously reported¹². In the case of denatured, synthetic α -chains, dicyanide haemin was added to α -globin before mixing with native CO β -chains. The resultant tetramer was purified and characterized as previously reported for β -chain mutants, and unpublished equilibrium data showed P_{50} and n values indistinguishable from native HbA (J.T., J. Pagnier, D. Shih and K.N., unpublished results). Further proof of the integrity of reatured α -subunits is shown above where O₂, CO, and methyl isocyanide dissociation rate constants for the α - and β -subunits in four haemoglobin samples are presented. These parameters were obtained by replacing O₂ with CO, CO with NO, and methyl isocyanide with CO and using the analyses described by Olson¹⁴⁻¹⁶. All time courses were fitted to two exponential expressions with equal spectral amplitudes at the wavelengths of maximum absorbance change (usually 420 nm). Gibson, Olson, and co-workers¹⁴⁻¹⁶ have shown previously that the faster components for O₂ and methyl isocyanide dissociation correspond to ligand displacement from β -chains. This assignment is confirmed by the α (Gly-E7)₂ β (His-E7)₂ data because for this protein the unaltered rate constant corresponds to the larger dissociation rate constant assigned to β -chains in native and *E. coli* wild-type haemoglobin. The association rate constants shown in Table 2 were assigned in this way. Wild-type and mutant sperm whale myoglobins contain an N-terminal methionine⁹. However, the ligand-binding, optical absorbance and electron paramagnetic resonance properties of the synthetic, wild-type protein are virtually identical to those of the native protein purchased from Sigma (B.A.S., K.D.E., S.G.S., R.J.R., A.J.M. and J.S.O., unpublished results). These observations are supported by the similarities of the corresponding dissociation rate constants in Table 2. Finally, similar wild-type control experiments were carried out for the ligand association reactions, and the synthetic *E. coli* His-E7 proteins had very similar rate parameters ($\pm 50\%$) to those of the native protein. Mb, myoglobin.

indirectly by complicated, model-dependent numerical analysis¹³. Even with this restriction, we had to resolve observed time courses into fast and slow reacting components which then had to be assigned to the α - and β -chains. This problem has been addressed previously for methyl isocyanide and O₂ dissociation reactions of haemoglobin¹⁴⁻¹⁶. Our new data for the α -chain E7 mutants confirmed these earlier assignments (Table 1) and allowed others to be made for O₂ and CO reactions. In native R-state haemoglobin, the larger association and dissociation rate constants can be assigned to β -chains and the smaller constants to α -chains (A.J.M. *et al.*, manuscript in preparation). A complete summary of the rate and equilibrium constants for both haemoglobin and myoglobin is presented in Table 2.

As shown in Table 2, substitution of Gly for His-E7 in myoglobin and human α -chains caused a marked decrease in

§ Present address: Laboratoire de Chimie et Biochimie, Pharmacologiques et Toxicologiques, CNRS UA 400, 45 Rue des Saints-Pères, 75270 Paris, Cedex 06, France.

Table 2 Rate and equilibrium constants for O₂, CO and methyl isocyanide binding to native (His-E7) and mutant (Gly-E7) sperm whale myoglobin and human α - and β -chains in R-state haemoglobin at 20 °C

Protein	O ₂			CO			Methyl isocyanide		
	$(\times 10^{-6} \text{ M}^{-1} \text{ s}^{-1})$	(s^{-1})	$(\times 10^{-6} \text{ M}^{-1})$	$(\times 10^{-6} \text{ M}^{-1} \text{ s}^{-1})$	(s^{-1})	$(\times 10^{-6} \text{ M}^{-1})$	$(\times 10^{-6} \text{ M}^{-1} \text{ s}^{-1})$	(s^{-1})	$(\times 10^{-6} \text{ M}^{-1})$
Mb(E7 His)	14 ± 3	12 ± 2	1.2 ± 0.3	0.51 ± 0.06	0.019 ± 0.005	27 ± 8	0.12 ± 0.02	4.3 ± 0.3	0.028 ± 0.005
Mb(E7 Gly)	140	1,600	0.087	5.8	0.038	150	10	6.3	1.6
α(E7 His)	29 ± 5	10.1 ± 4.6	2.9 ± 1.4	3.2 ± 0.4	0.0047 ± 0.005	680 ± 110	0.14 ± 0.06	1.8 ± 0.1	0.078 ± 0.033
α(E7 Gly)	220	620	0.4	19	0.0067	2,900	0.21	0.28	0.75
β(E7 His)	100 ± 24	21 ± 6	5.0 ± 2.0	9.8 ± 1.9	0.0086 ± 0.0030	1,100 ± 500	0.57 ± 0.13	6.4 ± 0.8	0.089 ± 0.023
β(E7 Gly)	100	37	3.0	5.0	0.0130	390	0.67	4.2	0.16

All reactions were carried out in 0.1 M K phosphate, pH 7.0 for myoglobin (Mb) or 0.1 M Bis-Tris, 0.1 M KCl, pH 7.0 for haemoglobin (α - and β -mutants). The exact protein concentration varied: 2–10 μM haem concentrations were used in stopped-flow, rapid mixing and conventional photolysis experiments, whereas 20–50 μM haem concentrations were used in laser photolysis experiments. Under these conditions liganded or R-state haemoglobin does dissociate partially into dimers; however, all previous work indicates that the functional properties of the α - and β -chains in dimers are essentially identical to those in R-state, high-affinity tetramers^{14–16}. The dissociation rate constants (k) were measured by stopped-flow techniques as described in Table 1. Association constants (k') for ligand binding to R-state haemoglobin were determined from partial photolysis experiments. Only time courses showing <10% photodissociation were considered to ensure that $\text{Hb}_4\text{X}_3 + \text{X} \rightarrow \text{Hb}_4\text{X}_4$ was the only tetrameric reaction being measured. The methyl isocyanide and CO complexes were photolysed by a square-wave, 0.5-ms light pulse whereas the O₂ derivatives were photolysed by a 500-nm dye laser pulse. The parameters for the His-E7 proteins were obtained from multiple experiments (>4), and the errors represent standard deviations from the mean. The constants for the *E. coli* wild-type proteins were averaged with those for the native proteins as systematic differences were not observed. Fewer experiments were done with the individual Gly-E7 derivatives, but the errors for the His-E7 parameters are assumed to apply to the mutant rate and equilibrium constants. The equilibrium constants (K) were calculated from the ratio of the rate constants (k'/k).

oxygen affinity due to a ~ 100 -fold increase in the O₂ dissociation rate constant. This mutation also resulted in a five-fold increase in CO affinity for these two proteins, due mainly to an increase in the association rate constant. There were also marked increases (10–60-fold) in methyl isocyanide affinity for the Gly-E7 derivatives of myoglobin and α -chains, with different kinetic effects in the two proteins. In myoglobin the His-E7 to Gly mutation caused a ~ 100 -fold increase in the association rate constant for isonitrile binding with little effect on the dissociation rate constant, whereas the same mutation in α -chains had no effect on the isonitrile association rate but decreased the dissociation rate about 10-fold. In contrast, the His-E7 to Gly substitution produced little or no effect on any of the ligand-binding properties of β -chains in R-state haemoglobin (Table 2, last two rows).

Neutron diffraction studies have demonstrated the presence of a hydrogen bond between His-E7 and the oxygen ligand in oxy-myoglobin¹⁷; no such hydrogen bond is found in CO-myoglobin¹⁸. The high resolution X-ray crystallographic structure of oxy-haemoglobin suggests that there is a similar hydrogen bond in the α -subunits but not in the β -subunits⁵. Shaanan has pointed out in detail the similarities of the distal pockets in oxygen-liganded myoglobin and R-state α -subunits. He also concluded that β -subunits behave more like model haem compounds than either of the other proteins⁵. Our experiments complement this structural work and clearly demonstrate different functional contributions of the distal histidine to ligand binding in these globins. In both myoglobin and α -subunits, His-E7 stabilizes bound O₂ by about $-1.4 \text{ kcal mol}^{-1}$, presumably by hydrogen bonding, and sterically hinders CO binding by about $1.0 \text{ kcal mol}^{-1}$. The differences between the geometry of O₂ and CO bound to haemoglobin and myoglobin have been discussed elsewhere^{18,19}. In the case of methyl isocyanide binding, the steric hindrance by His-E7 in myoglobin is much larger ($\sim 2.0 \text{ kcal mol}^{-1}$) than it is in α -chains ($\sim 1.0 \text{ kcal mol}^{-1}$), in agreement with Shaanan's observation of less hindered distal pockets in both subunits of oxy-haemoglobin. The insensitivity of oxygen affinity to the His-E7 to Gly substitution in R-state β -chains is also consistent with earlier calculations indicating unlimited freedom of rotation of bound O₂ and the apparent lack of a hydrogen bond with the distal histidine in the β -subunit of oxy-haemoglobin⁵.

We thank Drs Perutz, Dodson and Shaanan for their helpful comments. This work was supported by the MRC, the NIH (grants to K.N., S.G.S., J.S.O. and R.J.R.) and the Robert A. Welch Foundation (J.S.O.). J-P.R. was the recipient of an EMBO long-term Fellowship.

Received 14 July; accepted 30 September 1988.

- Pauling, L. *Nature* **203**, 182–183 (1964).
- Collman, J. P. *Accs Chem. Res.* **10**, 265–273 (1977).
- Moffat, K., Deatherage, J. R. & Seyberg, D. W. *Science* **206**, 1035–1042 (1979).
- Phillips, S. E. V. *J. molec. Biol.* **142**, 531–554 (1980).
- Shaanan, B. *J. molec. Biol.* **171**, 31–59 (1983).
- Mims, M. P., Porras, H. G., Olson, J. S., Noble, R. W. & Peterson, J. A. *J. biol. Chem.* **258**, 14219–14232 (1983).
- Nagai, K. *et al.* *Nature* **329**, 858–860 (1987).
- Perutz, M. F. in *Molecular Basis of Blood Diseases* (eds Stamatoyannopoulos, G., Nienhaus, A. W., Leder, P. & Majerus, P. W.) 127–178 (Saunders, Philadelphia, 1987).
- Springer, B. A. & Sligar, S. G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8961–8965 (1987).
- Nagai, K. & Thøgersen, H. C. *Nature* **309**, 810–812 (1984).
- Nagai, K. & Thøgersen, H. C. *Meth. Enzym.* **153**, 461–481 (1987).
- Nagai, K., Perutz, M. F. & Poyart, C. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7252–7255 (1985).
- Sawicki, C. A. & Gibson, Q. H. *J. biol. Chem.* **252**, 7538–7547 (1977).
- Olson, J. S. *Meth. Enzym.* **76**, 631–651 (1981).
- Olson, J. S., Anderson, M. E. & Gibson, Q. H. *J. biol. Chem.* **246**, 5919–5923 (1971).
- Reisberg, P. I. & Olson, J. S. *J. biol. Chem.* **255**, 4151–4158 (1980).
- Phillips, S. E. V. & Schoenborn, B. P. *Nature* **292**, 81–82 (1981).
- Hanson, J. C. & Schoenborn, B. P. *J. molec. Biol.* **153**, 117–146 (1981).
- Baldwin, J. J. *J. molec. Biol.* **136**, 103–128 (1980).

Corrigenda

Evidence from cathodoluminescence for non-volcanic origin of shocked quartz at the Cretaceous/Tertiary boundary

Michael R. Owen & Mark H. Anders

Nature **334**, 145–147 (1988).

WE regret that we failed to acknowledge that sample material was provided by Glen A. Izett, US Geological Survey.

Structure of antibody hypervariable loops reproduced by a conformational search algorithm

Robert E. Bruccoleri, Edgar Haber & Jiří Novotný

Nature **335**, 564–568 (1988).

THE reference to the crystallographic structure of the monoclonal antibody Fab fragment HyHEL-5 was inadvertently omitted. The reference reads as follows: S. Sheriff *et al.*, *Proc. natn. Acad. Sci. U.S.A.* **84**, 8075–8079; 1987. Also, in the legend to Fig. 3, the amino-acid residue H 100 was incorrectly specified as aspartate. This residue is, in fact, an asparagine.

The big Big Bang book

John Maddox

Stephen Hawking's A Brief History of Time has been the publishing phenomenon of the year. The book may have more buyers than readers — but does that matter?

THOSE who worry about the supposed public ignorance of science must surely be comforted to know that in the United States there are now in circulation 600,000 copies of Professor Stephen Hawking's book *A Brief History of Time*. For weeks now, the book has been in the best-seller lists, in Britain (where Bantam Press says it has 150,000 copies in print) as well as the United States. With Christmas on the way, sales could easily exceed 1,000,000 within a year of first publication. Is not this vast circulation of a tract (in the best sense) offering up-to-date if provisional answers to all reasonable questions about the origin, mechanism and fate of the Universe proof of public hunger for enlightenment and a sign that the hunger has been substantially satisfied?

Sad to say, most of those to whom I have been putting questions like these during the past few days seem not persuaded that Hawking's book is a sign that public indifference towards science (and the hostility that indifference breeds) is nearing its end. Indeed, and this may be sadder still, the most common reaction is best described as professional competitiveness, or otherwise as envy: people look abstracted as they multiply \$18.50 by 600,000 and then take 10 or 15 per cent of that.

Curiously, among roughly a score of people I have questioned during a visit to California (not all of them scientists) I found none who did not know of the book, three who owned a copy and none who had yet read it. This seems odd for a volume of only 198 pages whose author's estimate of the reading-time can be inferred, from his statement that 1,000 calories of nutriment will be required to capture its information content, as roughly half a day.

Indeed, there is a strange embarrassment about the book. People say it is a 'cult' book, or describe Professor Hawking as a 'cult' figure. In California, well-used to the coming and going of gurus differing in persuasions and persuasiveness, this explanation may seem natural. But even California cannot have absorbed all those 600,000 copies. The plain truth is that Professor Hawking's way of life and his reputation have gripped the public imagination in a way that is almost as phenomenal as the sales of his book.

The details are well-known and are retold in passing in the text. Hawking

developed motor neuron disease while still a graduate student, and has been in a wheel-chair ever since. His capacity for speech, already seriously impaired for 20 years, was entirely destroyed four years ago by a tracheotomy, with the result that he now communicates by means of a word processor linked to a speech synthesizer.



Sales talk — Bantam's poster for the British edition of Hawking's book.

During this long period, he has been extraordinarily productive of innovation in gravitation theory and quantum gravity, taking it as a guiding principle that his disability will not prevent him behaving as would any other creative scientist. One measure of his courage is that he is as likely to be found at meetings in Moscow or Texas as in Cambridge (England), where he holds Newton's chair in mathematics.

Hawking's hold on the public imagination is exemplified by the tale told last week by a Californian Englishman who explained that a chance acquaintance at a motor-service station, learning that he was a scientist, had asked "Do you know Professor Hawking? He's my hero". But an academic architect, whose trust in science is not much more robust than that of a rabbit in a fox, saw Hawking's hold on his students as a "Strangelove complex". "This", he said, "is what scientists want to be like, brains driving speech synthesizers".

There is no doubt that the vividness of

these images has been engendered by the publicity Hawking personally has given to his book. A few months ago, for example, he made a tour of campuses in the United States, everywhere speaking to crowded auditoria. But now the more envious of Hawking's colleagues are saying that his publishers are exploiting his disabilities commercially, which is the silliest canard. People who write books have an obligation (not merely to their publishers, but to themselves) to ensure wide circulation.

The more serious question is that of what the influence of the book will be. It is remarkable that there should seem to be so many buyers and so few who acknowledge being readers. Is it possible that the book itself is not as appealing as its title? Sadly, re-reading the book I still have the impression that it does not live up to its promise to explain the whole of recent cosmology (inflation, string theory and all that) to the ordinary person willing to make some effort.

My disappointment — shared, I see, by Professor Owen Gingerich (p.288) — is all the greater because Hawking's text begins so well. He teases the reader with questions that should spring to mind. Why did not people think that the Universe is expanding (or contracting) until the 1920s? What would the Universe be like if Newton's law of gravitation were not an inverse-square law? Special relativity is splendidly explained, but there are many among Hawking's general readership who will ask whether the old twin-paradox (that the returning traveller ages less than his stay-at-home twin) is completely exorcised by the remark that there is no such thing as absolute time. The first dozen or so pages of the text have a freshness that soon melts away as the complexity of the ideas is magnified with the passage of time. I doubt that even the most diligent readers, for example, will know what particle spin means if they did not have an inkling before embarking on the book.

Yet Hawking's chief message does indeed shine through: the gravitation theory of general relativity is so modified by quantum theory that the literal singularity that occasions the Big Bang would not have occurred, the behaviour of the Universe would not necessarily have been discontinuous at the time supposed to be its beginning and, as a result, one can speculate about what went before. That, of course, is a message fit for a guru. It is also an extremely important message for us all. Perhaps it will not matter if only a fraction of the 600,000 books in circulation are read cover-to-cover if the suspicion begins to form in people's minds that the Big Bang in its strict interpretation is just another fairy story. □

John Maddox is Editor of Nature.

Through a magic casement

John Cairns

What Mad Pursuit: A Personal View of Scientific Discovery. By Francis Crick. *Sloan Foundation Science Series/Basic Books*: 1988. Pp. 182. \$16.95. To be published in Britain in Spring 1989 by Weidenfeld & Nicolson.

CRICK's autobiography starts off as the straightforward story of an intelligent boy, brought up in a Protestant environment, who finds himself drawn to atheism, and thence to physics, in the face of his family's unquestioning religious beliefs.

When the war ended in 1945 he was 29 years old and, as a physicist, had successfully designed acoustic mines that were believed to have sunk many hundreds of ships — hardly, one would think, the ideal preparation for a Nobel prize in medicine. The government wanted him to continue designing weapons, but the curiosity of the agnostic was a stronger force and Crick decided that the most interesting subjects were "the borderline between the living and the nonliving, and the workings of the brain". At this point, he received a lot of advice from such a distinguished list of people that the reader comes to realize that Crick must have done remarkably well in his wartime work at the Admiralty, though this is something he modestly glosses over. After receiving the experts' conflicting opinions, he spent two years fruitlessly studying the physical properties of cytoplasm. Finally, in his thirties, he went to join a new Medical Research Council unit, set up in Cambridge for the study of protein structure by X-ray diffraction.

Now, at last, his career was underway, and the tale moves from one excitement to the next — from protein structure to DNA structure, and from there to the coding of protein synthesis. It is, by now, a well-known story. The double helix is the hero, of course, and what we are being given is yet another version of the leaf-fringed legend that haunts about its shape (for, if Crick can choose to take his title from Keats's *Ode to a Grecian Urn*, I too should be allowed to rummage about in the same poem). Finally, the sedges of Cambridge wither, and he retreats to California and the Salk Institute to wrestle with his other chosen subject — the brain.

That is the simple synopsis. But the book is much more than an autobiography. Its primary concern is with the place of ideas and theory in the biological sciences. Watson and Crick were impelled forward by an Idea. Crick points out that, even before they saw the meaning of the structure of DNA in 1953, Watson had coined

the phrase "DNA makes RNA makes protein"; the magnitude of that leap in understanding will become clear if you look at a contemporary review of protein synthesis such as Borsook's article in the 1953 volume of *Advances in Protein Chemistry*. Even as they were wrestling with the structure of DNA, Watson and Crick were starting to wonder how *base*



A youthful Francis Crick — given time to think.

sequence could be translated into *amino acid sequence*; to see this in its proper context, you should remember that when Itano and Pauling described the electrophoretic abnormality of sickle cell haemoglobin, in 1947, they concluded simply that the mutant gene was "through some series of reactions... introducing the modification into the hemoglobin molecule that distinguished sickle cell anemia hemoglobin from the normal protein".

Once the structure of DNA became clear, Crick's attention turned to the coding problem and he was led to propose the Adaptor Hypothesis, several years in advance of the actual discovery of transfer RNA. Before any part of the genetic code had been deciphered, Crick and Brenner had shown it was a commaless, non-overlapping triplet code, Brenner had demonstrated that there were chain-terminating signals, and the two of them had deduced the existence of messenger RNA.

Much of the book is devoted to the important part played by speculation in the history of molecular biology. This

aspect was largely missing from Watson's two best-sellers on the subject — *The Molecular Biology of the Gene* and *The Double Helix*; in the first, Watson gave a dazzling account of what might be called the inventorial view of molecular biology, and in the second he wrote almost exclusively about the personalities and scandals and the quest for fame and fortune. Crick describes what now most interests him about those few years in the 1950s, when a small group found themselves in the happy position of being the first to think usefully about the storage and control of biological information. It was one of those rare moments when there is a revolution in our understanding of the world around us.

In terms of the history of science, the sequence of thoughts behind each of the discoveries was as interesting as the actual discoveries themselves. Crick points out that most of the ideas they had about the coding problem proved to be wrong, and this leads him into a discussion of the continuous, delicate interplay between hypotheses and facts — what class of hypothesis is needed for each particular occasion, and how one should decide which facts are important and which are misleading. He goes on to describe what he thinks is the key difference between successful theories in biology and in physics. "Biologists must constantly keep in mind that what they see was not designed, but rather evolved. Physicists are all too apt... to concoct theoretical models [in biology] that are too neat, too powerful and too clean."

Here perhaps he could again have gone back to Keats, who wrote that the true mark of intelligence is an unwillingness to fill in the many gaps in our understanding with a set of irrational beliefs. Keats called it "negative capability". I wonder if that is not the essence of Crick's brilliance and clarity of thought. Certainly, it would account for his atheism and for his present distaste for much of what is called cognitive science.

Part of Crick's interest in that brief period of intense theorizing may relate to his current preoccupation with the brain and the problem of consciousness. The human brain finds things out by blundering around between various possible hypotheses. On certain rare occasions, a group of scientists suddenly find themselves looking, with a wild surmise, at some uncharted sea, and that is a time when their thoughts have to run far ahead of the facts. So perhaps there is something special to be learnt from the successes and failures of all those theories of the 1950s. Is Crick possibly thinking about a theory of theories, that might lead to another way of thinking about consciousness? ▶

Autobiography is a treacherous undertaking. The more you cover up, the more you are likely to reveal. Someone once wrote that he had never seen Francis Crick in a modest mood; it was a memorable phrase. Now, at long last, we are hearing from the man himself. Only in one respect do I think him misleading and immodest. He makes it all seem too easy. I have watched him at several Cold Spring Harbor symposia. He would sit in the front row, listen intently to every talk, interrogate, dominate, and end the week in a state of exhaustion. It seems for him there is nothing between rest and over-

drive. This capacity for prolonged and intense concentration is one side to his character that does not appear in the book.

Many readers will be struck by the thought that Crick belongs to a bygone age, when biologists were given time to think. What granting agency today would give several years of support to a young scientist who just wanted to build models? What 30 year old would now dare to embark on such a perilous pursuit? □

John Cairns is a Professor in the Department of Cancer Biology, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA.

"Amoebas at the start were not complex They tore themselves apart and started sex"

Mark Ridley

The Evolution of Complexity By Means of Natural Selection. By John Tyler Bonner. Princeton University Press:1988. Pp.260. Hbk \$40; pbk \$13.95.

WE DO NOT really know how many species are alive today. Until recently, the estimate was 'a few' million, a rough extrapolation from the fact that taxonomists have described about a million modern species. The latest estimate is a factor of ten higher, at about $3-4 \times 10^7$, the increase having been inspired by detailed sampling of the beetles in tropical trees.

Therefore, between the origin of life and the modern variety of forms, there must have been more than 30 million different lineages. It is difficult to estimate how many more. George Gaylord Simpson once guessed that 4×10^9 species have lived throughout evolutionary history. His estimate should probably now be revised upwards; but many of those species would have shared the same lineages, so the actual number of lineages must be between the total number of species that have ever lived and the total number alive today. Maybe 10^9 is a reasonable guess. Now, one *but only one* of those lineages runs through a long series of extinct forms from the origin of life to ourselves. A sure method of making many biologists hopping mad is to identify that particular lineage with the 'main line' of evolution or even with evolution itself.

Bonner is not one of those biologists. He is quite happy to talk about the evolutionary 'progression', of 'evolution from small to large' and 'evolution from simple to complex'. He is not interested in all the lineages in which evolution has gone from simple to simple, small to small (for the bacteria are still with us), or even from complex to simple, or large to small. All

patterns can be found among the 10^9 lineages available. I therefore found myself repeatedly disagreeing with Bonner. The disagreement, fortunately, is mainly about words. Neither his argument, nor his evidence, concern 'evolution' as a whole; and if every time he says 'evolution' we substitute 'evolution down the one lineage in 10^9 that happens to run from the origin of life to humans beings' it usually makes equally good grammatical, and better biological, sense.

Bonner's first subject is size. From the origin of life onwards, the size of the largest kind of animal or plant present at any one time has increased. There were only unicellular organisms 3×10^9 years ago: now there are blue whales, and there were dinosaurs not so long ago. Bonner explains the trend by what he calls ecological 'pioneering'. If a niche in nature is unexploited, a pioneer ecological species will soon invade it. Likewise, over evolutionary time, species will tend to evolve into places where there is reduced competition. The niches for small organisms have been filled ever since the origin of life, and progressively larger forms have been able to pioneer the niches for larger and larger types. The process, as Bonner discusses, would have been speeded up when sexual reproduction evolved — which explains my titular quotation from Arthur Guiterman. The argument is a form of Darwin's 'principle of divergence', though Bonner does not mention it.

He then shows, in what is the main original result in the book, that the histological differentiation of living things is proportional to their body size. Bonner has compiled estimates of the number of cell types in various kinds of organism. Small, unicellular organisms have one cell type, mushrooms have about seven, squids and Diptera about 55, and vertebrates more than 120. Actually, there is no increase in the average body size between the forms with seven cell types and those with more than 120; but the trend is clear between the end points.

The number of cell types is Bonner's main criterion of organismic complexity. It, like size, will tend to evolve higher

extreme forms by the principle of divergence, and Bonner argues that there are advantages in the division of labour too.

The book is not only about why more complex, and larger, forms have arisen later in evolution. Bonner discusses complexity in four areas of biology: genetics, embryology, ecology and animal behaviour. Behaviour is included as the pinnacle of complexity. Bonner likes analogies between principles in different fields, and he is interested, for example, in the abstract similarities between the development of behaviour and morphology. He does not, however, appear to have read J.W.S. Pringle's classic paper of almost 40 years ago (*Behaviour* 3, 90-110; 1951). Pringle discussed the analogy between learning and evolution, both more deeply and more exactly than Bonner does. Pringle defined complexity in terms of information theory, and noticed that it is characteristic of learning to increase the complexity of an organism's behaviour patterns. Like natural selection, learning increases the frequency of otherwise improbable events and builds up biological complexity as it does so.

Bonner has aimed to write a synthetic work. He draws on a wide range of evidence from the four areas of biology, and almost any biologist will learn something new. It is hardly fair to demand that another whole subject should have been included, but palaeontology is an obvious gap. Bonner has not ignored the fossil evidence, but he has made no use of the modern palaeobiological ideas on long-term evolutionary trends. He discusses trends in size and complexity purely in terms of natural selection, but (as palaeobiologists have pointed out) trends can also arise when a character is correlated with speciation, or extinction, rates. A graph in the book suggests that more complex species have higher extinction rates, although it is difficult to be sure how much of the effect is a taxonomic artefact. If more complex forms do have higher extinction rates, the force of 'species selection' will tend to *decrease* the level of complexity in biological systems.

Bonner has a gentle style. *The Evolution of Complexity* is written for the general biological reader and each concept, such as speciation, is explained in elementary terms when it first appears. He does not go in for any cheap showing off: he only discusses things he is prepared to take seriously, and does not caricature his opponents, or make spurious claims to originality. I doubt whether the book contains any important new insights, but it makes pleasant reading and may direct somebody to a new line of thought on this apparently timeless problem of evolution. □

Mark Ridley is in the Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

Government of the planet

Steven M. Stanley

The Ages of Gaia: A Biography of Our Living Earth. By James Lovelock. W.W. Norton/Oxford University Press: 1988. Pp. 252. \$16.95, £13.50.

JAMES Lovelock publicized his stimulating Gaia hypothesis during the 1970s, in part through collaboration with Lynn Margulis, an eminent cell biologist. The hypothesis asserts that life on Earth sustains habitable conditions for itself by actively regulating its environment on a global scale. In this world view, feedback from the abiotic environment stabilizes the global ecosystem.

Gaia was the Earth goddess of Greek antiquity. The Gaia hypothesis does not represent quite the apotheosis of our planet that its name, taken literally, would imply, but it does assert that the Earth, together with its life, is a kind of super-organism that maintains itself through a homeostasis akin to that of simpler forms of life. This elevated status does not go down well with many Earth scientists, who see organisms as modifying but not fully regulating their environment.

In this new portrayal of Gaia, Lovelock counters earlier criticisms that the Earth's interior is inorganic and beyond biological control by comparing it to the inner wood of a giant redwood. The simile is imperfect, however, in that the living tree synthesized its wood, whereas the mantle, core and much of the crust pre-date life on Earth. Furthermore, most definitions of life include not only the capacity for self-regulation and growth, but also for self-replication. Judgement of the Gaia hypothesis in these terms quickly degenerates into semantics, however, and life has never been amenable to proper definition in any case. Far more important is the veracity of the fundamental tenet of Gaia — that life, since shortly after its appearance, has in various ways sustained and improved its physico-chemical environment so as to flourish.

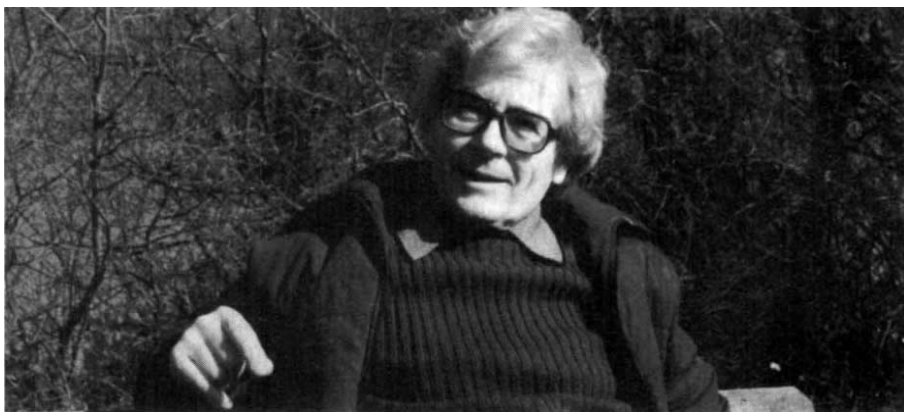
Although he makes his living as an inventor, James Lovelock was trained as a medical researcher. Thus, it was apt that he should coin the term 'geophysiology' for the discipline predicated upon Gaia. Here Lovelock notes an interesting historical parallel. James Hutton, the father of modern geology, was the first Western thinker to view the Earth's dynamism as constituting a kind of physiology, and Hutton, too, was trained as a physician. The overt attempt to establish a large new field of science almost singlehandedly is a rarity in the modern age, the other obvious example

being E. O. Wilson's launching of socio-biology — a field that, despite much controversy, has clearly taken hold. Lovelock's proposal for the legitimacy of geophysiology remains less substantial, which is not to say that none of his specific ideas will prove to have merit. His book should be read.

The Ages of Gaia takes its name from Lovelock's attempt to show how life has maintained a habitable Earth while evolving in the course of more than three billion years. In his view, Gaia awakened in Archaean time and then episodically shifted her homeostatic balances to new equilibria during biotic transformations. In this book, as in his earlier writings,

tive photosynthesizers to oxygen toxicity.

Lovelock offers no biotic reason for the build-up of oxygen to approximately its modern level by 600 million years ago, when animal life was diversifying rapidly, and his suggestions for biotic control beyond that point will seem far fetched to many readers. He adopts the conventional view that the oxygen flux into the atmosphere is governed by the rate of carbon burial — buried carbon being unable to recombine with oxygen — and that the oxygen flux from the atmosphere is governed largely by the rate of exposure of reduced material. He goes on to propose more unorthodox ideas about oxygen homeostasis, among them one in which



James Lovelock — the geophysiology with questions still to answer.

Lovelock concentrates on atmospheric chemistry and global climate. His initial formulation of the Gaia concept stemmed from his conclusion that life is responsible for maintaining a chemical disequilibrium in the Earth's atmosphere that is anomalous with respect to conditions on lifeless planets. Methane, for example, reacts with oxygen in sunlight and is sustained in the Earth's atmosphere only through ceaseless organic decay.

One of Lovelock's more imaginative sub-hypotheses is that, during the part of Precambrian time when oxygen sinks prevented photosynthetically produced oxygen from building up in the atmosphere, methane from bacterial decomposition rose to a relatively high atmospheric concentration. It then broke down in the upper atmosphere under the influence of ultraviolet radiation to form a smog of complex organic chemicals that strengthened the greenhouse effect, warming the Earth's surface. When oxygen finally began to build up in the atmosphere, it reacted with methane, lowering the concentration of this greenhouse gas and cooling the Earth. Lovelock suggests that this episode produced the ice age that is recorded in rocks about 2.3 billion years old. He suggests that the atmosphere then re-equilibrated within perhaps 200 million years, with oxygen remaining at a level well below that of the present day, the hypothetical limiting factor being the weak resistance of primi-

forest fires regulate the system, with intervals of heavy global burning accelerating carbon burial and causing atmospheric oxygen to increase. At such times, he argues, forests become even more flammable. The ensuing devastation results in a decline in carbon production and burial, which reduces the flux of oxygen to the atmosphere and thus reverses the trend.

Criticisms of Gaia have taken two forms: some people argue that the proposed regulatory mechanisms cannot or do not work, others that evolution could never have produced Gaia. The great complexity of the global ecosystem is at the heart of both classes of criticism. How can global regulation be attributed to a narrow set of controls, such as some of the effects of forest fires? How could life possibly control the salinity of the oceans? In addressing the other issue, how Gaia could develop in the first place, Lovelock emphasizes that success in natural selection is not simply a matter of flourishing within the environment, but of changing that environment to one's own advantage. This idea is applicable to individual populations, but how can it apply to a global ecosystem containing vast numbers of distinctive populations, each facing a unique set of selection pressures, with none able to benefit from its own minuscule contribution to such variables as global temperature or the level of atmospheric oxygen? ▶

And what of the great mass extinctions? If, as some experts estimate, 95 per cent of all species died out in the Permian mass extinction, would not a gaian system have died with them? Global biotic crises have generally occurred every 20 or 30 million years. If Gaia existed before each crisis, she failed during it, in the sense that abiotic causes of heavy extinction overrode her homeostatic mechanisms. The massive extinctions that ensued would

have damaged Gaia further, leading to still more extinctions, and so on. How could Gaia have survived such downward spirals? This present volume is imaginative and delightfully written, but full of unanswered questions. Perhaps Lovelock will address them in his next book. □

Steven M. Stanley is a Professor in the Department of Earth and Planetary Sciences, The Johns Hopkins University, Baltimore, Maryland 21218, USA.

The thin green line

Norman Myers

Public Policy and the Misuse of Forest Resources. Edited by Robert Repetto and Malcolm Gillis. *Cambridge University Press*: 1988. Pp. 432. Hbk \$69.50; pbk \$27.95. To be published in Britain in December, hbk £45; pbk £19.50.

DEFORESTATION of the tropics has been proceeding at an increasing rate since about 1970, yet virtually all analyses of the issue have failed to tackle the core factors at work. They have responded to the symptoms rather than the problems themselves. Still less have they addressed the sources of the problems, to be found in macro-level processes of policy making and planning.

Hence the value of this book — it is the first to engage in methodical assessment of the covert causes of deforestation, concentrating on “the impacts that governments, most of which are committed in principle to conservation and wise resource use, are themselves having on the forests under their stewardship through policies that inadvertently or intentionally aggravate [forest] losses”. Such policies range from those that deal specifically with forestry matters (like timber royalties and logging regulations) to those bearing on tax and trade regimes, industrial incentives, land-tenure laws, settlement programmes, agriculture, population, indeed development overall. The policies concerned were mainly instituted for worthy objectives, including industrial or agricultural growth, regional development and poverty alleviation. But all too often they “have resulted in economic and financial losses while contributing to the depletion of forest resources”. At the end of the book, those policies that can be altered to reduce forest loss without foregoing economic advancement are identified.

The overall argument is presented through case-study examples from Indonesia, Malaysia, the Philippines, China, Brazil and West Africa (also, for comparative purposes, the United States, where certain irrational policies still hold sway). In Indonesia, where timber exports have been earning gross income of

\$2 billion per year, inefficient economic policies in forestry industries (in the form of exceptional subsidies, tax exemptions and heavily subsidized credits among other types of protection) have not only fostered unsustainable logging; they have caused a loss of potential harvesting payments plus other revenues worth \$4.4 billion during the period 1979–1982 alone. In Brazil one of the main causes of deforestation has been cattle ranching: strongly promoted by all manner of incentives worth \$2.5 billion, it has proved to be a highly successful commercial activity for individual entrepreneurs, while imposing a net drain on the national economy. During forest clearance, ranchers have burned marketable timber worth \$5 billion, ostensibly on the grounds that they could make easier profits through government subsidies. Fortunately, in both Indonesia and Brazil, things have recently improved somewhat through direct intervention on the part of the countries' presidents.

The contributors of the 12 case studies amply document the ways in which forestry policies have frequently generated ‘timber booms’ and other scrambles for short-term profits. Yet the greater part of deforestation is attributable to non-forestry policies. Among them are excessive clearing of land for plantation crops such as rubber, palm oil and cacao, and for mining and roads, dams and other large infrastructure projects. Overall, the

principal form of policy-induced deforestation is through measures that impoverish households living close to the margin of subsistence, especially in rural areas: “These include pricing policies and investment priorities biased against the agricultural sector, development strategies that depress the demand for unskilled labor, and agricultural policies that favor large farms over smallholders. Such policies retard the demographic transition, make rural populations more dependent on natural forests for subsistence needs, and increase the concentration of agricultural landholdings”. It is primarily as a result of them that there has been a mass migration of small-scale cultivators into Brazilian Amazonia, culminating in the burning of some 80,000 square kilometres of forest during the 1987 dry season.

On top of all these mis-directed and mis-applied policies of the governments of the nations directly concerned, external factors have contributed to the general abuse of forest resources. For instance, trade barriers protect wood-processing industries in the developed world, thus usurping many of the benefits from forest-based industrialization that could accrue to developing countries. Much of the excessive logging has been carried out by multinational timber corporations based in the developed world.

Public Policy and the Misuse of Forest Resources is a pioneering appraisal of the entire policy context within which deforestation is progressively promoted. The editors are to be congratulated on their analytical acumen: the argument is pursued with detailed documentation throughout, and is rounded out with first-rate summary chapters which introduce and close the book. □

Norman Myers, Upper Meadow, Old Road, Headington, Oxford OX3 8SZ, UK, is a consultant on development and environmental matters. He was formerly on the staff of the United Nations Food and Agriculture Organization's Forestry Department.



“Protect the forest from fire” — on a billboard in Jagdalpur, India, a man is seen setting fire to a forest in this life and as a consequence is roasted over a fire in hell. From *The Hour of the Fox: Tropical Forests, the World Bank, and Indigenous People in Central India*, by R.S. Anderson and W. Huber. Publisher is University of Washington Press, price \$25.

Sticky answers to sticky questions

John Tyler Bonner

Topobiology: An Introduction to Molecular Embryology. By Gerald M. Edelman. *Basic Books: 1988. Pp. 240. \$21.95.*

THE ideas put forward by Gerald Edelman in his new book will certainly please developmental biologists, for they are entirely in tune with what is generally perceived to be the fundamental problem of development for many years to come. To begin with we all agree that at one end of the scale we must look at the genetics and, more particularly, at the molecular genetics of developmental processes. This is an area that is now advancing at a rapid pace. We also all agree (or at least those who have thought about it) that development is something that must be understood in an evolutionary context, for it is development that has evolved to produce adults of different structure.

Bridging the divide

Edelman bridges this divide in the following way. First he points out something that has been stressed by many people over the past few years: not all gene effects on development are simply the protein products of transcription and translation, but there are many secondary and tertiary effects, and ones even further removed from the initial proteins that play a vital role in moulding complex organisms. These involve chemical inductors, cell activities of many interesting sorts, including ultimately the activities of cells in the nervous system. Edelman thinks of cells as being the primary actors in carrying out the gene instructions, no matter how distant they may be from the cell actions. Cells can divide, move and die, and all of these processes play a key role. He is particularly interested in the substances that affect cell adhesion, an area where he has made valuable contributions, and shows how these cell adhesion molecules could carry out distant gene instructions to produce all sorts of tissue configurations including those we generally refer to as pattern. For pattern he also recognizes the importance of the reaction-diffusion models inspired by Alan Turing, but is especially attracted to the mechano-chemical models of George Oster, James Murray and Garrett Odell because they can directly include the fluid dynamic properties of the cell.

The book stresses the aristotelian, or perhaps better the waddingtonian notion of epigenesis: the sequence from gene product to ultimate form is the result of a series, or often a network, of causal molecular events which follow one

another. This route is, however, well canalized (to continue with Waddington's terminology), and the sequence, even if slightly perturbed, will come back to a consistent sequence and a consistent final form. Furthermore, the currently popular notion of developmental constraints is also stressed. Epigenetic sequences are severely restricted and tend to be conservative because of their past history. Any changes in new developments during the course of evolution are built on the foundation of the earlier ones. For instance, the vertebrate body plan has a common embryology despite a great variety in the form of the adults, as Karl Ernst von Baer made clear many years ago. For any large evolutionary shift (macroevolution) Edelman considers Ernst Haeckel's 'heterochrony' (where the sequence, or time of appearance, during development of different parts of the body can change) to be the explanation, again a position quite in keeping with modern views.

Towards the end of the book there is a discussion of how the nervous system might be formed, and how this might affect behaviour. Here there is an even greater distance between the immediate gene products and the ultimate construction. This is due, as Maxwell Cowan and others have stressed, to the idea that the ultimate connections of the brain are not specified in the genome, but are the result of the functioning of the neuron networks. There is solid support for this notion, not least of which is the evidence for massive cell death; only those cells that happen to be properly hooked up and are able to pass signals remain. Nervous systems and brains are remarkably plastic in both their development and their behaviour, again a feature that emphasizes the great distance generated by the epigenetic causal sequences from genes to a final functional structure. In this case the result is a brain capable of an enormous variety of behaviour not directly controlled by genes.

I have spent so much time giving the message from Edelman's book because I think it is an important one. In general, the ideas are not new — they are either ancient or modern — yet they are very much on the minds of many biologists today. But there are aspects of his excellent summary of this grand theme that are fresh and deserve praise, while there are others (largely involving their presentation) about which I am less enthusiastic.

The most exciting portion of the book (and fortunately it is a sizable one) deals with the glycoproteins that are responsible for the mutual adhesion of cells. Here, Edelman summarizes the work from his own laboratory, but also the work of others on proteins which connect cells to non-cellular matrices and gap junction and desmosome connections between

cells. He puts the experimental work into a larger setting which encompasses all of animal development, and does this in an admirable fashion for there is no doubt that cell adhesion plays a fundamental role in development. Most general discussions of epigenesis in the past (including my own) have stressed induction and the passage of chemical signals and responses. Others have appreciated the key role of adhesion in animal development. But I have not seen it put together, from gene to pattern, in such an insightful manner.

Tough reading

But all is not happy with this book. It is unfortunately not well written. For some perverse reason I am reminded of Alexander Woolcott's remark about the prose of the early mystery writer, Edgar Wallace; that it read as though it had been written on a sewing machine. This description hardly fits Edelman's book, but there is something equally wrong. After all, the story is a straightforward one, yet it is a hard book to read. Relatively minor irritants are the periodic sentences in italics, which reminded me of borrowing someone else's textbook and finding that the owner has underscored what he considers key passages; I kept on having to stop and read the italicized passages twice, for the first time the effect was as though someone had shouted at me through a bull horn and nothing registered.

A more serious problem is the use of acronyms and neologisms, all of which require learning a whole new set of symbols when plain English would have worked remarkably well. Take "topobiology", the very title of the book. Even after re-reading its definitions in different parts of the text I remain puzzled, for it seems to be a code word for some of the main ideas Edelman has to explain. The dozen or so acronyms are even more confusing. For instance Lewis Wolpert's 'positional information' becomes PI (not to be confused with 'principal investigator'). It is rather like reading a Russian novel when one constantly struggles to keep the characters straight. Yet I would not want to discourage the reader. For, despite its imperfections, I found *Topobiology* a first-rate contribution to developmental biology and believe that the effort needed to read it is well worth the while. □

John Tyler Bonner is a Professor in the Department of Biology, Princeton University, Princeton, New Jersey 08544, USA.

• Just published by Chapman and Hall/Van Nostrand Reinhold is the seventh edition of *Van Nostrand's Scientific Encyclopedia*, edited by Douglas M. Considine. The encyclopedia now comes in two volumes and contains over 7,000 entries. Price is £99, \$150 until the end of this year. £115, \$175 thereafter. For a review of the sixth edition see *Nature* 303, 450 (1983).

Flying into history

Richard Mabey

Biographies for Birdwatchers: The Lives of Those Commemorated in Western Palearctic Bird Names. By Barbara and Richard Mearns. *Academic*: 1988. Pp. 490. £17.50, \$35.

SHOULD birds, like dahlias and diseases, be named after people? It is a practice that has survived into the twentieth century, yet it seems a curious hangover from the age of magic, this occasional appropriation of a living creature for the person who first identified it, or shot it, or who just happened to be inspiring the namer at the time.

In this exhaustive study, Barbara and Richard Mearns have sketched the biographies of all 91 people who figure in the scientific or popular names of Western Palearctic birds. They are a motley bunch, chiefly from the nineteenth century. The clergy are strongly represented, as are army officers and colonial servants, for whom exploration was part of the job. Britain heads the list, with 29 subjects, followed by Germany with 25. Only three women feature, and one of these is the unfortunate Thekla Brehm, who died — quite uninterested in birds — when she was 24, and was commemorated in Thekla's lark by her grieving naturalist father.

It is a long book, intended to be used for reference. But enquirers within should be warned that they may not find kindred spirits behind their favourite species. The Northumberland engraver Thomas Bewick, who learned his craft by copying the inscriptions from gravestones, certainly deserved to have a swan named after him (by William Yarrell). But it is not so pleasant to read of the antics of the man commemorated in one of the Mediterranean's commonest warblers. Franco Bonelli (1784–1830) was a zoologist from Piedmont whose notebooks describe 146 methods of catching birds. He 'so admired' his eponymous warbler that he kept one in a cage to study its plumage changes. Colonel George Montagu, of Montagu's harrier, was no better, a great collector and slaughterer who nonetheless had the temerity to lament the havoc caused by this kind of behaviour in others.

Eleonora of Arborea (c. 1350–1404) sounds altogether more interesting and worthy. She was a reforming warrior princess, who had Eleonora's falcon named after her in honour of the legislation she passed protecting birds of prey. But the authors suggest that this was simply to guarantee a market in falcons for the nobility, and for a truly honourable and appropriate naming one has to go to the knot, the little wading bird that Camden



Mary Evans

Thomas Bewick — a deserving case.

thought (probably quite erroneously) commemorated King Canute's (or Cnut's) tideline attachments.

It is a fact of history that most nineteenth-century ornithologists were cold-blooded anatomists, with no great skill in field observation or love for birds, and it is no fault of the Mearns that it is mostly this unedifying aspect of our 'dominion over nature' that is immortalized in the names of birds. They have done a thorough and valuable job of research, even if the results do make them appear a little like Henry Tristram's (of Grackle fame) Griffon vultures, which "took a peculiar interest in taxidermy, scrutinising, head on one side, the whole operation of bird-skinning, and perfectly aware of the moment when a morsel would be ready, exhibiting a more than ordinary excitement when they saw the skin drawn back over the head". Personally I would have preferred the more critical response of the Ivory-billed woodpecker captured by Alexander Wilson (Storm petrel and phalarope) which drilled a hole the size of a fist in his wall and covered his room with plaster in its efforts to escape.

There was a parallel nineteenth-century tradition which deplored reckless collecting and favoured the study of live birds in the field, and it is a comment on past social attitudes that none of this school has been commemorated. The last new naming (the Corsican nut-hatch, dubbed *Sitta ledanti* after the Belgian ecologist Jean-Paul Ledant) was as recent as 1976, so there may yet be opportunities to mend our ways. I would like to nominate for future coins the still vacant titles of St Francis and St Cuthbert, Margaret Cavendish and John Clare. □

Richard Mabey, 10 Cedar Road, Berkhamsted, Hertfordshire HP4 2LA, UK, is author, among other books, of *Gilbert White: Biography of the Author of the Natural History of Selborne* (1986).

Observer observed

Robert M. May

Gilbert White and His Records: A Scientific Biography. By Paul G. M. Foster. *Christopher Helm*, London: 1988. Pp. 240. £19.95.

GILBERT White wrote only one book, *The Natural History and Antiquities of Selborne* (hereafter *Selborne*). The book was first published exactly 200 years ago, in 1788, and since that time new editions have continued to appear at an average rate of slightly more than one per year. The changing styles in the binding, illustration and design of the different editions form a record of changing taste and fashion against which the unchanging popularity of the book itself is remarkable.

Paul Foster is Head of English at the West Sussex Institute of Higher Educa-

tion. His book is not a conventional biography, but rather is an enquiry into the intellectual origins of *Selborne*: into how Gilbert White developed as an observer of the natural world around him, and how his book came into being. Foster, who appears to have an acquaintance with everything written by White (or written to him) that borders on the obsessive, surveys White's writing practices over more than four decades. The material includes: a *Garden-Kalender* which was kept as a record of the goings-on in the vicarage garden for over 20 years from 1751; the *Flora* of Selborne which records a concentrated study over the one year 1766; correspondence with a brother who sent White natural history specimens from Gibraltar while he was posted in the garrison there from 1769 to 1772; and the meticulous *Journals* of meteorological and other natural events which White kept from 1768 up to his death in 1793.

Foster presents this material in a lively

way, showing how the record of more-or-less private writing underpins *Selborne*, and "demonstrates White's systematic, if sometimes halting, development". More than this, Foster singles out those aspects of the work that set it apart from the usual compendium of curiosities, and give it a contemporary resonance.

One such aspect is White's use of deliberate experiments in pursuit of an understanding of what was going on around him. These vary from simple observations as to the efficacy of removing mouldiness from some melon plants by "the experiment of clipping off the infected parts with a pair of scissors", or experimenting with peat dust instead of old thatch in the bottom of potato trenches, through to sowing "a pot with cucumber-seeds, and set[ting] it by the parlour-fire for experiment-sake", to rather complex and detailed experiments carried out in connection with ventilating a hotbed. The great age of experiments in farming by the Russells and others (of which the Rothamsted Field Station is one legacy today) still lay about half a century ahead in White's time, and it would be interesting to see a larger analysis that explored the extent to which White's style and ideas, and this larger enterprise, were part of the same broad current.

Foster also presents an assessment of the accuracy of the observations recorded in White's writing, and in his *Journal* in particular. As Foster remarks, "to gain purchase on the level of accuracy deployed in the records of a natural historian two centuries ago is not easy. The best way would be to walk the lanes and paths of Selborne with White and to verify his

observations as he made them, but that is clearly impossible". Foster gives a careful analysis of the likelihood that White actually saw what he says he saw, and that the weather was as he recorded it, and decides if there are errors "they have crept into the record against White's practice, and against his conscious caution". To a degree unusual in the eighteenth century, White seems to have understood the procedures and problems of scientific record-keeping, noting that "the bane of our science is the comparing one animal to another by memory in order to describe it".

Although he spent most of his life in Selborne, White was not isolated. His correspondence includes exchanges with some of the most distinguished naturalists and travellers of his time (Lightfoot, Pennant and Banks among them). He maintained a lifelong involvement with the affairs of the Oxford college (Oriell) where he was an undergraduate and later Fellow, although — as described by Foster — the running of the college at that time seems to me to reflect little credit on anyone associated with it.

Foster's book is probably more for the *Selborne* aficionado than the general reader. It is, however, far more than the sadly common shoebox of file cards arranged by chronology or theme, and those with an interest in Gilbert White will find in it an engaging account of the intellectual odyssey of that remarkable man. □

Robert M. May is a Royal Society Research Professor in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PU, UK, and at Imperial College, London.

Radiating curiosity

Eric Denton

In Search of Nautilus. By Peter Douglas Ward. *Simon & Schuster: 1988. Pp. 238. \$19.95.*

THE pearly nautilus is one of the most beautiful and famous of all sea creatures, giving its name to the wondrous craft imagined by Jules Verne and to the submarine built early in the last century by Robert Fulton. It is evidently related to the fossil nautiloids and ammonoids and seems, much more than the coelocanth, to be the archetype of the 'fossil' animal. Peter Ward had two general aims in writing a book about this extraordinary animal. First, using nautilus research as an example, to show how fitful the scientific process can be. Secondly, to tell the story of nautilus itself. He has succeeded in both.

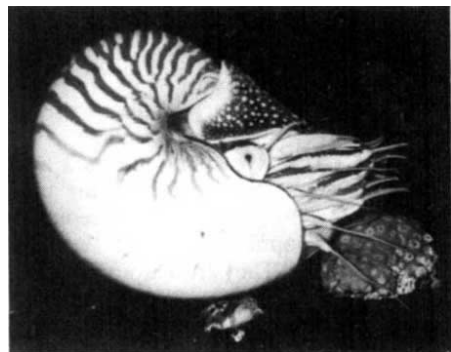
Ward begins by reviewing the earlier descriptions and speculations of Aristotle,

Robert Hooke, Richard Owen and others to bring his story to the end of the nineteenth century. The main business of zoologists was then widely believed to be that of understanding and following the processes of evolution, and great hopes were pinned on the study of the early development of animals during which they were thought to recapitulate their evolutionary history. The nautilus was felt to be a key animal. With this in mind, a very able Cambridge zoologist, Arthur Willey, spent the years 1894–1897 working on the nautilus in the south Pacific, always under difficult and sometimes dangerous conditions. He discovered a great deal but, although many of his animals laid eggs, he sadly failed to find any that were fertile.

The account of Willey's adventures and results is followed by an assessment of the researches of Anna Bidder, John Gilpin-Brown and myself in the early 1960s. My only 'complaint' about this account is that I am sure that I used either the expression, "damn all" or "sweet Fanny Adams" rather than that attributed to me on page

93. Ward then recounts the rapid advances made, mainly by American zoologists and geologists, during the past 20 years. He keeps a good balance between the development of scientific ideas, the ways in which the problems posed were solved, and the difficulties of research and dangers to researchers from boats, storms, sharks and bureaucrats. He is well acquainted with the scientists who have worked on the nautilus recently and has taken the trouble to meet those whom he did not know well.

The book concentrates on four sets of



Ancient design — the chambered nautilus.

problems. First, how does the nautilus achieve and use its buoyancy system? Secondly, where and how does it live? Thirdly, how many species are there and what is their place in evolutionary history? And, lastly, how does nautilus reproduce and what can we learn about evolution of cephalopods from its embryological 'history'? On all of these points Ward has a great deal of successful research to report, research that has given conclusive results which will not be overturned in future. He ends on two contrasting thoughts — that the latest work seems to show that, surprisingly, nautilus is a very young form in the midst of a great radiation of species and that, less surprisingly, its future is most threatened by man's predations.

A few years ago the New York Academy of Sciences decided to launch a project to help scientists and science writers to get books published for a popular audience. The aim was to communicate the excitement of discovery without drifting into superficiality. *In Search of Nautilus*, published as part of this project, is admirably successful in achieving the Academy's aims. It should attract a wide general readership, as well as the attention of professional geologists and zoologists who will find much of interest, lucidly and amusingly expressed by an acknowledged international authority. □

Sir Eric Denton, Fairfield House, St Germans, Saltash, Cornwall PL12 5LS, UK, is an Emeritus Professor of the University of Bristol and was Director of the Marine Biological Association's Plymouth Laboratory from 1974 to 1987.

Those not-so-dumb animals

Wayne M. Getz

The Honey Bee. By James L. Gould and Carol Grant Gould. W. H. Freeman: 1988. Pp. 239. £14.95, \$32.95.

HONEY bees are the *sapiens* of invertebrate evolution, or so we are led to believe by James and Carol Gould. Some myrmecologists may not agree with them; but, whatever your persuasion, you will be convinced after reading this book that honey bees are remarkable organisms, living in an even more remarkable society.

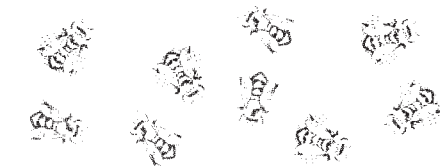
Like its companion volumes in the Scientific American Library, *The Honey Bee* is printed on high-quality paper and includes beautiful colour pictures and illustrations. A small line drawing of a worker bee is cleverly placed on the right-hand corner of consecutive obverse pages, so that in flipping through the book from back to front the reader is treated to a cinematic illusion of a dancing bee.

The first chapter is a succinct but charmingly written history of beekeeping, the second a brief introduction for the uninitiated to the life of the honey bee. The

remaining chapters are woven around the particular research interests of the first author, James Gould, an ethologist at Princeton University. Thus, the title is misleading, because the book primarily deals with communication, navigation and learning. But this choice of topics is appropriate because the book complements rather than competes with other recent accounts of the honey bee, notably T.D. Seeley's *Honeybee Ecology: A Study of Adaptation to Life* (Princeton University Press, 1985) and M. Winston's *The Biology of the Honey Bee* (Harvard University Press, 1987). The Goulds, however, underplay the importance of olfactory communication in mediating the organization of honey bee colonies; and they ignore a number of fascin-

rely on vision, ultraviolet radiation, polarized light and magnetic fields for determining flight orientation. There is also an intriguing exposition of the honey bee dance and the controversy, which brewed in the late 1960s and early 1970s, over Nobel laureate Karl von Frisch's interpretation of the dance as a communication system. Unfortunately, one section is marred by an excessively long description of how, as a student, James Gould — using some cleverly and some not so cleverly designed experiments — was able to vindicate von Frisch.

As explained in the book, variations in the dance constitute a language that enables returning workers to communicate with great precision the location of nectar, pollen, water and propolis, as well



ating sociobiological issues which relate, for example, to brood care and kin recognition.

On the whole, the writing is absorbing — especially in the section where the reader is guided through the wonders of a four-layered hierarchy of honey bee navigational systems which respectively

as sites for bivouacked swarms to establish new hives. What, though, can we deduce about the intelligence of honey bees from studying their ability to communicate spatial information through this dance? The Goulds raise the possibility that a bee behaves as though it has a mental map of the area in which it forages and is able to reference this map while performing the dance. As they point out, this level of intellectual reasoning in a mere bee is heresy in our hubristic world of smart human beings and dumb animals.

The second half of the book is devoted to making a case for some spark of intelligence in the bee with its one-million-neuron brain. But, on the penultimate page, the Goulds get cold feet and draw a parallel with chess-playing machines that can delude us into believing they are intelligent when, indeed, we know that these computers are only electrons whizzing around in circuits. (Or is that how humans think as well?)

This final prudence is understandable, but it is not in keeping with other parts of the book where controversial hypotheses are stated as fact. For example, I would like to see the evidence that "a single cell in the olfactory lobe can encode an odor". I also reserve judgement on the comment that "far and away the most complex communication system in nature other than our own is found in a mere bee" until we have decoded, among other things, the song of the humpback whale.

For all that, *The Honey Bee* has much to commend it. In particular, it provides a highly readable account of why this insect is considered by some to be the cleverest creature in the invertebrate world. □

Wayne M. Getz is a Professor in the Department of Entomology, University of California, Berkeley, California 94720, USA.

Kastanienbaum

Beneath the spreading chestnut tree
we study late and long
the humble Myxophyceae
in systems right or wrong.

They thrive in mud or watershed —
non-saline or marine
and vary from a purple-red
to shades of bluish-green.

Their only chlorophyll is *a*.
They flourish in the light.
They fix their CO₂ by day
and nitrogen by night.

Some cells are sheathed or capsulate;
some lack all forms of dress,
but, sexless, do not hesitate
to show their nakedness.

We must attempt to settle now,
but cannot soon decide
on what they should be called, and how
they should be classified.

With varied principles at stake
we argue endlessly
around the bench, beside the lake,
beneath the chestnut tree.

From *The Biology of Algae and Diverse Other Verses* by Ralph A. Lewin, an expanded version of his original *The Biology of Algae and Other Verses* published by Boxwood, price \$12.50.

In the Faculty Club Lounge

He totters to his chair, an aged don
Whom critics, in their time, once scowled upon.

On issues past, his fantasies and state
Were subjects of derision and debate . . .
But now, no more.

Subsiding in his seat
He takes the weight of ages off his feet.
His spectacles slip wryly from his nose.
He turns from retrospection to repose,
And contemplates his fellows from a haze

Of old cigars and even older baize.
His effigy, in marble and in bronze,
Unrecognized by later crops of dons,
Discreetly gathers dust beside the stairs.
And, from that landing too, surveys his heirs

Who used his textbooks once, though
now no more
Except to prop an argument, or door.
They know him only, in this latter age,
By petty footnotes on a yellowed page.

His faculties and focus all but gone,
He daily, dimly, donly doddles on.

Reviews on IMMUNOASSAY TECHNOLOGY

Edited by S. B. Pal

Volume 1.

October 1988 £37.50 256pp

234x156mm 0 333 452852

Apolipoprotein quantitation by ELISA: Technical aspects and clinical applications - J. B. Bury & M. Y. Rosseneu

Recent progress in the quantitation of phagocytosis - T. W. Jungi

Monoclonal antibodies directed against human immunoglobulins: preparation and evaluation procedures - J. J. Haaijman et al

Conjugation of haptens and macromolecules to phycoerythrin for application in fluorescence immunoassay - N. Monji & A. Castro

Nicotinic enzyme immunoassay - A. Castro & N. Monji

Applications of human monoclonal antibodies in non-isotopic immunoassays - S. A. Gaffan and M. C. Glassy

Adaptation of enzyme labelled immunoassay for measurement of biosynthesis of human - L. Morrison et al

Immunofluorescent methods for the assay of cytoskeleton antibodies in human sera - D. Zauli et al

The use of Western blot procedures in the analysis of *Herpes simplex* virus proteins - J. Kuhn et al

Non-isotopic studies of the TSH receptor - N. R. Farid and G. Fahraeus-Van Ree

Index

Volume 2.

October 1988 £37.50 216pp

234x156mm 0 333 46849X

Human monoclonal antibodies produced by Epstein-Barr virus immortalized cell lines: technical and theoretical principles - M. Steinitz

Sensitive enzyme and erythro immunoassays for detection of human pregnancy - S. K. Gupta

Automated methods to test polymorphonuclear leucocyte function: - P. De Paoli & G. F. Santini

The Epstein-Barr virus hybridoma technique for production of human monoclonal antibodies - R. Burioni et al

Applications of liposomes in non-isotopic immunoassays: a review - N. Monji et al

Establishment of a particle concentration fluorescence immunoassay for the measurement of human IgG in culture supernatants - M. B. Compas & Y. G. Alevy

Establishment of a very sensitive avidin-biotin ELISA for the measurement of human IgG - Y. G. Alevy & C. M. Blynn

Methodology, limitations and uses of antigen-binding assays in the analysis of T-cell receptor-antigen interactions - B. E. Elliott et al

Spin immunoassays - C. C. Curtain & L. M. Gordon

Enzyme-linked immunoassays for detection of anti-drug antibodies - J. W. Coleman

IqE measurement in serum and culture supernatant - C. Gaveriaux & F. Looor

Index

Order these books through your bookseller or send your order to:

Richard Gedye, Macmillan Press Ltd,
Houndmills, Basingstoke RG21 2XS, UK.

What was where when

Christopher Chippindale

Past Worlds: The Times Atlas of Archaeology. General editor Chris Scarre. *Times Books, London/Hammond, Maplewood, New Jersey: 1988. Pp.319. £29.50, \$85.*

Time and space are the natural dimensions against which the variables in human history and prehistory are measured, so an atlas of period-by-period maps is a convenient way to present the pattern of sites of human habitation or activity. *Past Worlds* is the newest in this genre, offering 122 double-page spreads in a very large format; six sections cover the world from hominoid ancestors to the industrial revolution. The colour maps are drawn with shaded relief, and give a vivid impression of the set of the landscape.

Running across this landscape are, of course, the distributions of individual recorded sites — but individual sites are less and less the stuff of field archaeology, which is why distribution maps are a declining species. The trouble lies in the vagaries of site creation and of site preservation and discovery.

Some lifestyles, particularly mobile ones, do not create defined sites, but rather a diffuse scatter of material across the landscape. Others, like a tradition of building in mud-brick, generate massive mounds of debris. Accidents of preservation bury some sites beyond reach, or long ago exposed them to destruction. Elsewhere, conditions have both protected sites over the long term, and now make them accessible to study. Again, the patchy distribution of Pleistocene coastal sites follows from the postglacial rise in sea-level; because sites have survived only when they were raised on cliffs, or in those few places where uplift has raised the old shoreline, the bald pattern of distribution is not directly indicative of where Pleistocene settlement actually was on the coast.

Equally, the pattern of recorded sites measures the pattern of fieldwork; as the joke goes, there are no distribution maps of archaeology, only distribution maps of archaeologists. So it is that the blank regions on such maps may indicate blanks of site formation, blanks of site survival or blanks of field-survey — as well as real empty spaces in an ancient time.

The advent of better dating methods has meant that time has become more controlled as a variable, with the analytical units of Stone-Bronze-Iron Ages and their subdivisions giving way to absolute chronologies. In their place have come a variety of divisions by period according to a variety of criteria, so there is no longer

such a standard set of conventional, distinct periods.

Faced with the reduced usefulness of simple period-maps of spot sites, Scarre and his large team of contributors have gone in for very full explication. There are 14 introductory double-page spreads of method and analysis, and a map occupies only a part of each main spread. The rest is given over to pictures of artefacts, reconstruction drawings, case-studies and explanatory text, so the book is less an atlas, more a pictorial encyclopaedia of archaeology with especially thorough maps. A number of spreads have no maps at all.

The authors of the ancillary material have generally coped well with their difficult task of explaining complex questions in a few pictures and a few hundred words, but much less space is left for specific information about what was where when. Because the maps only occupy part of each spread, often less than half, they have to be smaller and show less than full-page maps would do. The largest North American map, for example, is too little to provide state boundaries, and it is not possible to use the atlas to find the main sites in, say, Minnesota. Some important sites fall in the gaps between the chosen topics. Mont Bego, the second region of later prehistoric art in the Alps, appears on the map of Bronze Age Europe (though under its old name of Monte Bego and wrongly placed in Italy), but the much more important region of Valcamonica does not appear, perhaps because its art spans several periods as well as the Bronze Age. While some things are lost, others are promoted to prominence — a few single sites have a whole spread to themselves, and feudal Yorkshire is given as much room as the Islamic world. Some users will think that the loss in factual detail is not worth the benefit of all the ancillary explanation, handsomely illustrated though it is.

The book is generally sound on contentious issues, although Americanists may be surprised by the judgements made about controversial early sites, but it will soon be out of date in regions such as Australasia where discovery is proceeding apace. Yet many of the maps draw together a remarkable amount of information from scattered sources; they must have been a real labour. The production, colour-printing and index of sites are good. This will be a useful volume, as well as handsome on the coffee table. □

Christopher Chippindale is Editor of Antiquity and Assistant Curator in the Cambridge University Museum of Archaeology and Anthropology, Downing Street, Cambridge CB2 3DZ, UK.

• The next review supplements to appear in *Nature* are Textbooks (which will be published in the issue of 9 March 1989) and Spring Books (issue of 20 April).

Selection among the systematists

Philip Kitcher

Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science. By David L. Hull. University of Chicago Press: 1988. Pp. 586. \$39.95. To be published in Britain in December, £31.95.

ACCORDING to Legend, the white coats that scientists wear are emblematic, for scientists are surely the good guys, serious, humble, scrupulously honest and dedicated to sacrificing themselves in the pursuit of Knowledge. Legend has not gone unchallenged in the two decades since James Watson offered his famous account of the race to find the structure of DNA, but it lives on in journalistic reports of scientific practice, in the reflections of many scientists, and in the descriptions offered by many historians and philosophers of science. David Hull believes that it is time to take a closer look. His fascinating book attempts to portray the factors that are really at work in the process of science, to present one case in detail and to put forward a general theory.

Using the vocabulary of Legend, we might divide the book naturally into two parts: data and hypotheses. For his data, Hull has gone primarily to the community (or, more exactly, communities) of systematists whose doings he has followed closely for about 15 years. He offers a vivid chronicle of the formation of research groups, of coalitions and alliances, of collaboration and confrontation. The drama often runs high. Friendships (even marriages) are made, hostility and jealousy thrive, and, from time to time, the heroes plot revenge against those who have tried to thwart their ambitions. Yet, like any good primatologist, Hull has great sympathy for his subjects. They emerge from his narrative without white coats, not always admirable, but thoroughly human.

Hull advances hypotheses at two levels. In the early chapters he uses the data to support claims about the kinds of motives that are prominent in the daily practice of science. This sociological theorizing is to be embedded in a more general analysis, advanced in the second half of the book, in which science is viewed as a selection process. In any generation of science, ideas are spawned in profusion. Some of these ideas, perhaps because of their own intrinsic merits, perhaps because of the activities of the scientists who promulgate them, are more likely to be replicated than others. While the less fortunate vanish without trace, the successful ones find their ways into the brains of many scientists, perhaps even securing the status of textbook embodiment. Scientists themselves are active in the war of ideas, for each is driven not only by curiosity (which Hull

recognizes as an important part of scientific motivation) but also by the desire to secure credit. So we are offered the picture of a community of investigators who sometimes cooperate and who offer one another recognition for their contributions, but who compete for the flourishing of their own intellectual property. Success comes in having your ideas widely used and properly attributed to you.

"Legend-bashers typically attract science-bashers, and Hull can expect some of his readers to conclude that he has revealed science as brutal warfare among ambitious egoists, a process that has nothing to do with rational argument or the uncovering of truth."

Objections to evolutionary accounts of human knowledge (and to darwinian accounts in particular) are familiar, and Hull attempts to defuse them by offering a very abstract analysis of selection processes. The core of his idea, familiar from his earlier writings, is that there are two important kinds of entities, *replicators* and *interactors*. Replicators are things that have the power to transmit their structure with reasonably high fidelity. Interactors cause differential replication of replicators by their interaction with some environment. *Selection* occurs when variation in the success of interactors causes the differential perpetuation of associated replicators. Starting from this general proposal, Hull surveys the variety of ways in which the natural selection of organisms occurs and tries to show that worries about applying selectionist ideas to the dynamics of knowledge are unfounded.

Presented with a gripping tale of science in the raw, some sociological generalizations about spurs to research and an abstract account of the workings of selection, it is natural to ask a number of questions. How typical is the case that Hull describes? How well does the case exemplify the general theory? How different is the perspective provided by the theory from that offered us by Legend? For all its many virtues, *Science as a Process* fails to resolve the doubts that lurk behind these questions.

Hull is sensitive to the charge that his quarrelsome systematists are atypical. He responds by alluding to other historical

and sociological studies and by denying the legitimacy of asking for the behaviour of the 'typical scientist'. In my judgement, these responses do not take full account of the force of the objection. Many social practices, including the academic disciplines of philosophy and literary criticism, would have served equally well to illustrate the power of the motives that Hull lays bare — philosophers, in particular, are a notoriously argumentative lot. Yet journalists rarely celebrate the progress of philosophy. If philosophy and science are subserved by the same mechanisms, what accounts for the difference in their apparent results?

That question is bound to loom large when one looks more closely at the issues in dispute between Hull's warring systematists. Those issues concern the 'philosophy' of taxonomy, the 'first principles' on which a 'proper' systematics should be based. The Legend-lover now understands why the behaviour Hull records was to be expected: of course, when scientists wander off into philosophical speculation there will be no objective standards to guide them and they will lose themselves in sterile wrangling.

Even if we set on one side concerns about the representativeness of the scientists whom Hull has studied, the relation of data to hypotheses still proves troublesome. Although Hull's chronicle is keyed to his sociological generalizations, it is simply not clear what aspects of the behaviour of the systematists are illuminated by the very general account of the selection process that he gives in the second half of the book. In part, this is a consequence of the lack of detail: Hull has very little to say about the processes through which scientific ideas are replicated or what makes them successful in the struggle for replication. But, in his defence, we could note that Darwin was forced to complain that his own ignorance about the processes of variation and inheritance was "profound". Perhaps we can then excuse the absence of detail in Hull's account, trusting that the specifics can be filled out later. Indeed — but in the interim we need a reason to think that a selectionist theory of the growth of science will do some explanatory work. Darwin succeeded, despite his professed ignorance, because he could fill the *Origin* with descriptions of phenomena that could be understood on the basis of common descent or of natural selection and which were otherwise inexplicable. *Science as a Process* contains nothing like that.

One facet of science that Hull regards as needing explanation is the fact that failure to acknowledge the ideas of another scientist is typically regarded as less serious than scientific fraud. His explanation turns on the idea that the theft of an idea only harms one person, while the

publication of fraudulent data can disrupt the research of many. Even if this is correct — and one of the obvious rival explanations is that judgements about failure to assign credit are often tempered by the suspicion that the borrowing was inadvertent — it is far from clear that Hull's evolutionary account has any role to play in articulating the explanation. The lacuna here points to an important flaw in the book. There are some interesting findings about the behaviour of a group of scientists and a very abstract account of the metaphysics of selection. The wheels, however, never quite mesh.

Legend-bashers typically attract science-bashers, and Hull can expect some of his readers to conclude that he has revealed science as brutal warfare among ambitious egoists, a process that has nothing to do with rational argument or the uncovering of truth. Yet the author himself has no truck with fashionable relativisms. He is confident that science makes progress, and that, as a way of attaining knowledge of the world "it beats all other ways hollow" (p.26). Indeed, at his most venturesome, Hull promises to show that science advances because of the polemics, the keen desire for revenge, the intellectual competitiveness, all the marks of the behaviour of his subjects that make defenders of Legend avert their eyes in horror. The promise envisages a thoroughly naturalistic approach to science in which it is shown how certain kinds of intellectual and social forces contribute to the attainment of epistemic goals. But Hull's book does not deliver. Nor could it, for it treats the transmission and origination of scientific ideas at far too abstract and imprecise a level, and it is self-consciously silent on some of the epistemological questions that any successful development of the naturalistic project should address.

Legend is overdue for replacement, and an adequate replacement must attend to the process of science as carefully as Hull has done. I share his vision of a serious account of the social and intellectual dynamics of science that will avoid both the rosy blur of Legend and the facile charms of relativism. The faults of his book are those of omission, but it is only fair to recognize that they are dwarfed by the accomplishments. Because of the author's deep concern with the ways in which research is actually done, *Science as a Process* begins an important project in the study of science. It is one of a distinguished series of books, which Hull himself edits, and which in recent years has greatly advanced the history, philosophy and sociology of science. The editor should be more than happy with the author's contribution. □

Philip Kitcher is a Professor in the Department of Philosophy, University of California, San Diego, La Jolla, California 92093, USA.

The better way to better health

Stephen Lock

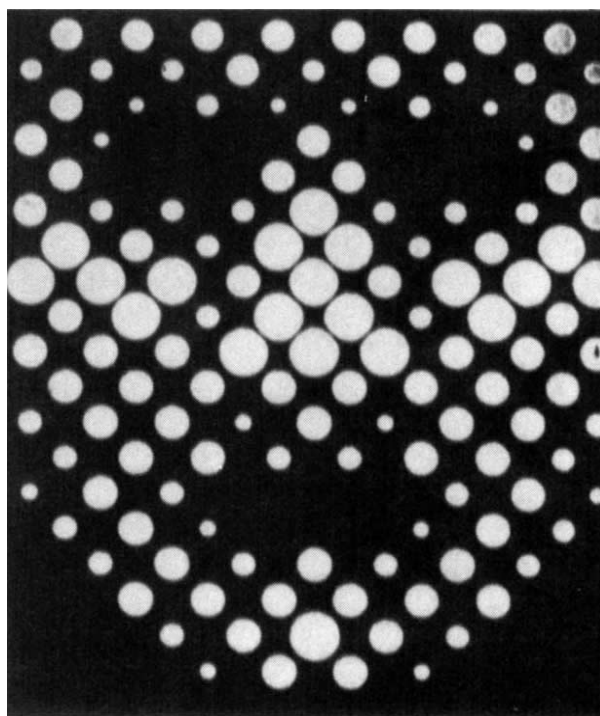
The Origins of Human Disease. By Thomas McKeown. Basil Blackwell:1988. Pp.233. £22.50, \$34.95.

UNDERPINNING today's craze for jogging and bran is a formidable amount of epidemiology, here cogently summarized in the late Thomas McKeown's last book. Uniting his principal interests in the history of populations and of infections with a forensic power of argument, McKeown concludes that three influences on health stand out: infection, malnutrition and population growth. Medical advances have played only a minor part in improving things, a conclusion that will hardly endear him to the medical establishment any more than did his earlier, similar claims. Nor will politicians welcome his emphasis that we already have much of the knowledge needed to improve health rapidly; what is lacking is effective management and a political commitment to equality of opportunities for good health. But McKeown is equal to any adversary and the blurb is accurate in describing his book as a *tour de force*.

Recent attacks on health care have been on different grounds: the economic incompetence of its delivery. But only a decade ago (when McKeown came on the scene) the intellectual wrangling was

about the role of modern medicine, with allegations that it was positively deleterious to people's general well being. To be sure, medicine has seen off some powerful critics before — Voltaire and Bernard Shaw among them — and in the late 1970s and early 1980s it had little difficulty in coping with two of the new ones. Ivan Illich, an "intellectual Tyl Eulenspiegel", as the *New England Journal of Medicine* described him, was opposed to any aspect of industrialized society, and his appeal to restore pain and suffering to the individual evoked little enthusiasm. Ian Kennedy, the 1980 Reith lecturer, argued with a lawyer's subtlety but his lectures were condemned for their superficiality, particularly his accusation that doctors had usurped control of ethical decisions.

Both Illich and Kennedy provoked a useful debate about what part medicine played in health. Nevertheless, intellectual rigour was added when McKeown joined in, because he had both the appropriate knowledge and the debating skills. A Canadian Rhodes scholar, he had held academic appointments in the basic sciences before going to Birmingham and becoming a foundation professor of social medicine at the age of 32. At Birmingham he created virtually a new subject, which came to concentrate on the changes in human populations and disease patterns. In some quarters his 1976 Rock Carling lecture on the role of medicine was interpreted as an attack on clinical care, and in a second edition, admitting that he had underestimated its value, he devoted



Arresting image — the World Health Organization's multi-language poster, produced in 1978, announcing a \$1,000 reward for the first person reporting a confirmed case of smallpox acquired by person-to-person transmission. The reward remained valid until the global eradication of smallpox was certified in December 1979. The picture is taken from *The Smallpox Story in Words and Pictures*, by Abbas M. Behbehani, published by the University of Kansas Medical Center, price \$15.95 (North America), \$16.95 (elsewhere). For a review of WHO's Smallpox and its Eradication see Nature 332, 755 (1988).

a new chapter to the topic. He did not recant, however, but persisted in his conclusion — that because clinicians saw the body as a machine, whose disorders could be put right in the same way as a car repair, they had overlooked the enormous contribution from modifying disease origins. Moreover, because the role of some influences had been over-emphasized (particularly medical care), society had misused its resources and needed to reconsider how such resources were allocated.

In his new book, completed just before he died in June of this year, McKeown classifies human existence into three phases. In the first (hunter-gathering) apart from accidents and tribal wars the important restraints on population growth were infanticide and malnutrition/starvation. The food supply increased 5,000 years ago during the second phase (agriculture), but again the benefits were outstripped by population growth, and malnutrition recurred. And then another predominant cause of illness and death arrived: infectious diseases. Probably originating from animals (measles from dogs, the common cold from horses, tuberculosis from wild cattle, plague from gerbils and influenza from pigs), they could become established and amplified not only against the background of malnutrition but because people now lived together in larger communities, under cramped conditions with poor hygiene.

In the West, the third phase (industrialization) began three centuries ago. Health standards and nutrition dramatically improved, while the population rose and has continued to do so — staggeringly so for the whole world (750 million in 1750, two (American) billion in 1930, five billion today, with probable stabilization at ten billion in 2100). In McKeown's view the early increase in Western populations owed everything to better nutrition, which increased resistance to the infectious diseases, the increase in the late 1800s also being linked to better hygiene (piped water, sewage and safe milk). Further, communities came to limit their reproduction using contraception, and infanticide disappeared once most children were wanted. Significant medical advances arrived only in the twentieth century, particularly emergency surgery, chemotherapy for puerperal fever and tuberculosis, and immunization against diphtheria and poliomyelitis.

For the individual these and other later developments were obviously important — but, for populations, their role was minor compared with that of better nutrition and bathrooms. Now, however, modern life has produced its flip side: the non-communicable diseases, which, given that human genes have remained unchanged and that these illnesses were

absent in wild animals, hunter-gatherers and agriculturalists, McKeown calls diseases of maladaptation. These diseases originate in the environment, through changes in diet (too little fibre and too much food, salt, fat and sugar); the use of alcohol, tobacco and illicit drugs; lack of exercise; and pollution.

One can argue with some features of McKeown's thesis: he is uncritical, for example, about 'bran-wagon' evangelism that explains everything from coronaries to varicose veins as being the consequence of too little dietary fibre. Similarly he surely underplays the value of modern drugs, such as the antibiotics, in prolonging useful and fulfilling life. But in global terms he is right. In future centuries even the West's current spectre, AIDS, will figure as only a blip on the population screen, comparable with the 'minor' effects of bubonic plague or the influenza pandemic — how could we conclude differently in a world where five million children under the age of five still die from diarrhoeal diseases every year?

But the great merit of McKeown's book is that it points to solutions within our grasp. If sufficient food was available to every citizen in the underdeveloped world, health and self-limitation of the population would follow. This approach has worked in China, Sri Lanka and Kerala, where, through the political will to attain equality in health, Western standards have been achieved without improving water quality, sanitation or personal health services (including immunization). As N.H. Antia wrote recently (*Nature* 335, 584; 1988): "Vaccines . . . should not divert [Indian] attention from the major problem — the need for egalitarian socioeconomic development that alone can eradicate the root cause, namely poverty".

The developed world should recognize that diseases, including the non-communicable ones, are not inevitable. Action is indeed being taken, from the top downwards. Abolishing cigarettes would eliminate one in three deaths from cancer and smoking is fast becoming unacceptable. The same will happen with excessive alcohol consumption, and in eliminating the 3lb of chemicals we eat in our food every year. Nevertheless, the developed world has not abolished diseases of poverty; the greatest contribution to health would come from ridding our society of inequalities, which, as the Black Report showed, are still a blot on most countries. And, as McKeown himself wrote some ten years ago: "If a choice must be made, free school meals are more important for the health of poor children than immunization programmes, and both are more effective than hospital beds". □

Stephen Lock is Editor of the British Medical Journal, BMA House, Tavistock Square, London WC1H 9JR, UK.

Man and microbe

Christopher Lawrence

Robert Koch: A Life in Medicine and Bacteriology. By Thomas D. Brock. Science Tech Publishers, 701 Ridge Street, Madison, Wisconsin: 1988. Pp.365. \$35. Distributed outside North America by Springer-Verlag, £17.

UNTIL now there has been no book-length biography in English of the German bacteriologist Robert Koch. Indeed, as the author of the present volume notes, "even the German language biographies of Koch are dated and mostly inaccessible" (p.vii). This biographical void is perhaps partly related to Koch's authoritarian personality and his less than impeccable marital life when measured by the standards of the later nineteenth century. Nonetheless it is still surprising because Koch's life contains nearly all of the elements a scientific biographer might wish for: an earth-shattering discovery by a young, unknown and solitary doctor, a rapid rise to fame (and shame), a medical revolution and more than a hint of sexual intrigue.

Koch was born in lower Saxony in 1843. After a conventional medical education and an unremarkable bourgeois marriage he settled down in 1872 as District Medical Officer in the town of Wollstein, now in Poland. The 1870s were a turbulent decade in European medicine. The advocates of laboratory science were just beginning to turn the tide against the bedside men, and the proponents of a germ theory of disease were making their voices heard against the defenders of miasma or atmospheric pollution. These two controversies were related. By 1900 bacteria seen through the microscope on the laboratory bench had become the symbol of a new and scientific medicine.

Koch worked in an area where anthrax was rife. For a general practitioner, he was exceptionally well read in the medical literature and, more surprisingly, he was an exceptionally devoted experimentalist. After many months of painstaking research Koch convinced himself that anthrax was caused by a living organism and, more important, that this organism could survive as spores in the soil and in animal hides. Convincing himself was one thing, however. What about the world?

In April 1876 the provincial doctor took his preparations to the illustrious botanist Ferdinand Cohn and the great pathologist Julius Cohnheim at the University of Breslau. Both men were sympathetic to the germ theory and immediately championed Koch's work. From this point Koch entered a much larger public arena. He moved to Berlin and over the next ten years he and his collaborators created many of today's basic bacteriological tech-

Metastatic Spread of Cancer

G.V. Sherbet

This book outlines a new approach to the analysis of the mechanisms of metastasis by comparing the glioma, an intracranial tumour which ordinarily does not metastasize, to other metastasizing tumour models. This has produced a coherent picture of the operation of oncogenes and growth factors in the initiation, development and progression of neoplasms.

This is an essential book for those researching metastasis and will be of interest to all research oncologists.

1987 £45.00 275pp
234x156mm 0 333 443 624

Contents

Part 1

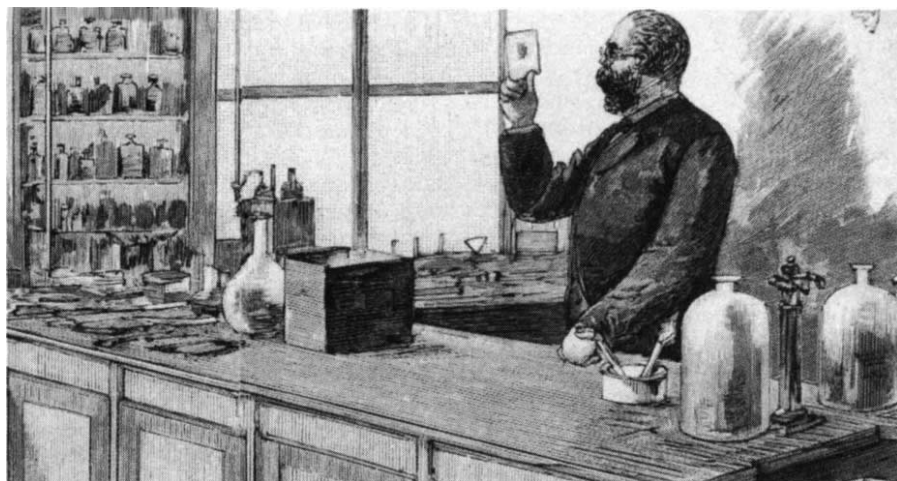
Introduction
The Intrinsic Malignancy of Gliomas
The Metastatic Cascade

Part 2

Heterogeneity of Tumours
Oncogenes in Tumour Development and Metastasis
The Genetic Instability of Neoplasms
The Generation of the Metastatic Phenotype

Glossary
References
index

Order this book through your bookseller or send your order to:
Richard Gedy, Macmillan Press Ltd, Houndmills, Basingstoke RG21 2XS, UK.



Mary Evans

Robert Koch — proponent of a new and scientific medicine.

niques. Koch's greatest public moment came with his announcement of the discovery of the tubercule bacillus in 1882. In addition he described the cholera vibrio and it was his co-workers and students who identified the bacteria responsible for many of the killer and crippling diseases of the nineteenth century, such as diphtheria, gonorrhoea and plague.

Thomas Brock relates all of these matters, although most of them would merit a small monograph in themselves. The biography is of a conventional sort; the Koch presented here is the one who was known from previous smaller and less satisfactory accounts. Brock gives us Koch and his warts in some detail, including Koch's famous announcement of his discovery of tuberculin as a cure for TB, his disputes with Pasteur and various other European men of science, his domineering character and competitiveness, the collapse of his first marriage and his remarriage when aged 50 to a 20-year-old art student (with whom he had had an affair since she was 17). Brock is perhaps cunningly trying to evoke our sympathy here by juxtaposing a photograph of the (to us) stern and matronly Emmy Koch (his first wife) with one of a sylph-like Hedwig Frieberg as she was when she first met Koch.

Although it is over 350 pages, this is not a long book. The illustrations, the short chapters and the layout combine to make it less extensive than it appears at first sight. The substance of the text also fails to confirm the first impression that this is a thorough reassessment of Koch's life and work. Brock offers no re-evaluation of previous work, nor do current trends in the history of science enter his interpretation. For him, Koch's discoveries were unproblematic pictures of nature which, willy-nilly, the world was obligated to accept.

Such an approach allows Brock to duck issues which otherwise his material would force on him. Thus when one of Koch's co-workers swallows a culture and develops cholera it proves that the disease was caused by the vibrio. When Pettenkofer

swallows a similar culture with no effect it merely proves he was lucky and Koch was right anyway. Similarly, Koch is praised for patience and faith when persisting with seemingly unfruitful experiments, while his opponents are accused of stubbornness for doing the same thing. (Koch himself, however, is called pig-headed for sticking to his views on the non-communicability to humans of bovine tuberculosis.)

Nonetheless, the strengths of this enjoyable book outweigh its weaknesses. It is readable and not too demanding. Although the author is a microbiologist he does not get carried away with technical questions, and for the non-bacteriologically minded explains his material lucidly. He gives us something of Koch the man, he has read the German material and quotes from a number of letters. The account is also decorated by letters from visitors, such as Arthur Conan Doyle, to Koch's Berlin empire. Brock is at his best in setting Koch's technical innovations in context. Here, in spite of all the talk of discovery and breakthrough, he shows that the establishment of bacteriology had less to do with claims about causality than the capacity of bacteriologists to repeat practical everyday tasks and, with their technology, to demonstrate their account of the world.

There is one wider question, however, which is not touched on at all, or at least only to endorse conventional wisdom. This is Brock's view that "to a real extent we owe our current good health and longevity to discoveries made by Robert Koch and his school". Against this it might properly be argued that, since antiquity, society has known about the health-promoting value of fresh water, a varied diet, clean air, good housing, exercise and relaxation. The politics of implementing these for everybody, rather than the difficulties of discovering bacteria, might be said to have been the biggest obstacle to our Western well-being. □

Christopher Lawrence is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BP, UK.

Renown in theory

Steven Shapin

The Pasteurization of France. By Bruno Latour. Translated from the French by Alan Sheridan and John Law. *Harvard University Press: 1988. Pp.273. \$30. To be published in Britain on 9 December, £23.95.*

HISTORIANS will not turn to this text for a definitive account of Pasteur's research and its institutionalization. (A long-awaited book by G.L. Geison will soon perform that function.) Latour is a philosopher-sociologist, and he uses historical material solely to illustrate a general theory of the nature of science, society and the relations between them. Indeed, Latour describes this volume as a "companion book" to the précis of his theoretical framework in *Science in Action* (see *Nature* 326, 754; 1987). Latour's divergent disciplinary allegiances are reflected in this new book's two radically different sections — one which belongs to a recognizably empirical 'social historical' genre, and a shorter, oracularly aphoristic 'philosophical' part written in an idiom that might best be described as a mixture of late Wittgenstein and high *chutzpah*: " 'Everything is necessary' and 'everything is contingent' means the same thing — that is nothing"; "Knowledge does not exist — what would it be?"

In Latour's case, the audacity is justified. He is one of our least boring commentators on science and society. Everything he writes is provocative, important and worth the closest scrutiny. If he is right, all existing research into the nature of science and its social relations is wrong, and we had best start again with him as our guide. And if he is wrong, he is only strategically wrong, and we still benefit from the acuteness of his tactical analysis of scientific practice.

How did it come to pass that there is a street named after Pasteur in every town in France? More generally, how are we to account for the 'success' of any particular piece of science or technology? Is it sufficient to point to the genius of the individual Pasteur? No, that can't be right. That Pasteur was a 'genius' was precisely the outcome of the work of many others — those who seized upon his work and declared him a 'genius'. Ideas do not diffuse of their own accord. They must be fetched and carried if they are to move. Thus, Latour points to the essential role of the pre-existing "hygienist" movement, with its talk of "contagion" and "morbid spontaneity", and its sanitary practices diffusely directed towards countless sources of environmental danger. Pasteur sought to interest them in his bacteriological ideas, and they were indeed interested, because pasteurism gave them a

plan whereby they could mobilize their skills and concentrate their forces against just a few points through which environmental threats might pass.

Are we then to speak of the 'truth' or 'efficacy' of Pasteur's ideas if we want to explain the pasteurization of France? That also can't be right, Latour says, since it was *because* his ideas were seized upon by others and inserted into their routine practices that they came to be accounted 'true' and 'powerful'. Hygienists, seeing what they could do with Pasteur's ideas, immediately celebrated the certainty and validity of his experiments. But there were some, such as Koch and Peter, who resisted, pointing — with support from contemporary canons of scientific method — to the "hasty generalizations" Pasteur made from a few experiments. Pasteur's

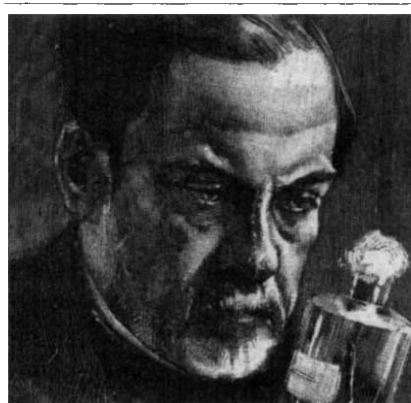


Image of intensity — the Pasteur of public fame. The picture is reproduced from the recent re-issue of René Dubos' Pasteur and Modern Science (Science Tech/Springer).

opponents were behaving no more 'unscientifically' than his supporters. Latour opposes realist and rationalist philosophies of science: 'truth' and 'power' are *outcomes* of what are usually called social processes; they can't be regarded as causes of scientific judgements.

France was not pasteurized through the inherent genius of the individual Pasteur nor because of the inherent truth of his ideas. Pasteurism was victorious insofar as its ideas and practices enlisted and were enlisted by numerous and powerful allies and insofar as they were made indispensable to others. Latour incisively analyses the processes by which pasteurians recruited their allies and inserted themselves into others' networks. Did farmers wish to prevent their cows from dying of anthrax, *vignerons* and brewers to make more reliable products, surgeons to heal their patients? Then they must acknowledge the existence of myriad microbes acting on them and their work. Who spoke for these invisible microbes? Pasteur and his colleagues. Those who wished to defeat the malevolent action of the microbes were obliged to pass through Pasteur's laboratory to accomplish their goals. The institutionalization of knowl-

edge and practice is seen as the constitution of a new "obligatory point of passage" through which others must pass to do their work. Pasteur's laboratory was the place where his skills were most effective. By funnelling others' interests through it, Pasteur himself grew strong. Latour's analysis of the pasteurian circulation between field and laboratory and back again is a brilliantly innovative way of conceiving the institutionalization of science.

Conversely, some groups resisted Pasteur because they sensed that his skills and practices conflicted with their professional interests. Ordinary medical practitioners, for example, initially saw in Pasteur's methods and doctrines a "negation of medical work". Bacteriological concepts struck at the root of the relationship between doctor and individual patient, just as pasteurian stress on preventative measures directed towards populations conflicted with family physicians' emphasis on the cure of individual patients. Until the shift from preventative vaccines to curative sera, the pasteurians had not found a way to enlist such physicians as allies. And the physicians, "fighting to save [their] profession" and "to advance their own interests", found "disputable" and unscientific what hygienists found indisputable (pp. 117–127). Evaluations of the truth and usefulness of scientific claims are informed by evaluations of their relationship to existing skills.

Summarized this way, Latour's analysis seems both highly persuasive and in no way incompatible with British and American sociological and social historical perspectives developed over the past ten to twenty years by such scholars as Barry Barnes, H.M. Collins, G.L. Geison, Andrew Pickering and Charles Rosenberg. Yet, puzzlingly, Latour is extremely rude about work in this genre. Neither social context nor social interests can be used to explain the natural sciences; "social explanation" is, he pronounces, simply "obsolete" (pp. 255–256). The apparent reason for this judgement points to the originality of Latour's position as well as to the problems of using that framework to say anything about more traditional perspectives. The pasteurization of France was the creation of a new society — now containing the microbes for whom Pasteur spoke — as well as a new science. In Latour's lexicon, "interests" and "society" are effects of scientific work, and cannot be used to explain that work.

Indeed, Latour says that we must no longer speak of science and society but only of "stronger and weaker associations". Nor may we decide in advance what these networks consist of, whether they are made up of human or non-human agents. Everything can be what he calls an "actant" — men, machines, microbes. That some are said to possess agency, will and activity and others not is the result of

attributional work in science. The distinction between men and the world on which they act is not to be a resource but a topic. Again, we cannot explain social action by referring to human beings as active and machines or bacteria as acted upon. This radically anti-reductionist and anti-realist stance arises from a French semiotic tradition that is neither well known in Anglophone scholarship nor adequately explained in this book. Is it a methodology and an analyst's conceit or is it an ontology? Is it a position about 'external reality' (as humble Anglo-Saxon empiricists think of it) or is it about the self-contained 'signifiers' of semiotic discourse?

One has to think that this semiotic perspective, when properly explained and understood, will resolve evident contradictions in Latour's work between programmatic injunctions and empirical practice. Thus, his abundant references to the professional "interests" of physicians are presumably not what they seem to be to Anglophone readers. And this new semiotic idiom will also explain how the use of such approved latourian locutions as "social groups", "networks" and

"strong and weak associations" is different in kind to that of supposedly banned explanatory entities such as "social context" or "social interests".

Not just historians, sociologists and philosophers, but also scientists and members of lay society, have traditionally been interested in the relations between entities they call science and society. Latour instructs us that we have not just been giving bad answers but asking bad questions. The price of assimilating Latour's approach is that those who do so can have nothing to say to those who continue to be concerned about what they see as 'science-society' relations. The links between the academic science studies community, on the one hand, and scientists and laity, on the other, will be well and truly broken. The radical originality and wit of Latour's approach is hugely attractive. But many scholars will require a better understanding of the foundations of that work and a better sense of where it leads before they decide to pay the price.

Steven Shapin is Reader in the Science Studies Unit, University of Edinburgh, 34 Buccleuch Place, Edinburgh EH8 9JT, UK.

Yummy, yummy?

John Yudkin

The Paleolithic Prescription: A Program of Diet & Exercise and a Design for Living. By S. Boyd Eaton, Marjorie Shostak and Melvin Konner. Harper & Row: 1988. Pp.306. \$17.95*.

MUCH has been written about the possible causes of the rise in prevalence of coronary heart disease, diabetes and some sorts of cancer, mostly in the wealthier countries but now also being seen in parts of the Third World. One source of clues for the change in disease pattern lies in the comparison of the life-styles of the populations of places where the incidence of these diseases is very different (such as North America and Western Europe, against the non-industrialized nations of areas of Africa, South America and Asia).

Such comparisons have been the main reason for the widespread belief that at least some of the diseases of affluence are caused by diets low in fibre, because the diets of many countries that are largely free from these diseases are fibre-rich. This approach makes the assumption that the ideal life-style of the human race, and one to which we should return, is that of people living in pre-industrialized countries. But it is difficult to sustain such an argument in view of their lower stature,

and of the high prevalence of infection and infestation, susceptibility to food shortages as well as to specific nutritional deficiencies, and altogether a low life expectation found in such countries.

How we live now in the 'advanced' countries has been achieved not by one revolution but by two. The more recent industrial revolution of the past 200 years or so has chiefly affected only the wealthy West. By far the greater and more important change was the Neolithic revolution of some 10,000 years ago, which occurred with the discovery of agriculture. Up to that time, mankind like all other species of animal had to get its food by hunting and gathering.

The Neolithic revolution, by which man learned to produce food by deliberately growing it, changed not only the sorts of foods that made up the human diet, but produced a whole new mode of existence: no longer nomadic, with a different and seasonal round of activity, and a social difference in the patterns of work and in the rearing of children. The few millennia that have elapsed since then have been far too short for there to have been any radical change in our genetic make-up; it still functions best with the sort of diet, activity and general personal relationships that were seen in our pre-Neolithic ancestors of Palaeolithic times.

The authors of *The Paleolithic Prescription*, two anthropologists and a radiologist, present precisely this thesis: that the best way of ensuring optimal health, including the avoidance of heart disease, obesity, diabetes and other diseases of affluence, is to adopt a Palaeolithic life-

style. They derive their evidence both from palaeontology and from the few tribes that still practise a largely hunting and gathering way of life. The authors elaborate their thesis in a splendidly readable, simple and lively style that seems most convincing.

It would be good to be able to add that the result is flawless exposition, but there are indeed flaws. Some of the suggestions are beside the point, such as avoiding oysters because they are rich in cholesterol; not many individuals regularly consume oysters as an important part of their diet. Some of the assertions are unproven and unlikely (that a low intake of fibre or of carotene can cause cancer), while others are quite unwarranted (that most people are short of calcium and that even those individuals that follow the "Paleolithic prescription" might be better off by taking vitamin supplements).

But the main flaw is the authors' failure to discuss those characteristics of human biology without which neither the Neolithic revolution nor the industrial revolution could have taken place. The invention of agriculture, with its revolutionary changes from a nomadic to a sedentary lifestyle, and from a diet rich in meat to one rich in starchy foods, was possible only because the human species is omnivorous, a quality that it shares with only a few other species such as the rat and the pig. The technological changes of the industrial revolution made it possible increasingly to produce new foods by separating the qualities of palatability from the qualities of nutritional value. Up to that time, it was true for all species that the foods they liked were the foods they needed: it is no longer true that, for human beings, palatability is an infallible guide to nutritional value.

It is also wrong to claim that the "Paleolithic prescription" is new, or even, as the publisher says, "groundbreaking". The idea has been around for at least 30 years; for example, it appeared in an article in *The Lancet* in 1956 and has been refined and elaborated in many later publications. We can now understand why our sugar consumption is so huge; it is not because, as the authors say, "developers of sugar beet and sugar cane plantations, together with industrialists, food technologists and advertisers, have encouraged" us to take more and more sugar-rich foods and drinks, but simply because we like them. Food manufacturers, technologists and others have devoted the same amount of energy and expertise to try and induce people to take food yeast; they have signally failed to do so simply because people don't like the taste of yeast. The omission of such considerations greatly detracts from the value of this book. □

John Yudkin, 20 Wellington Court, London NW8 9TA, UK, is Emeritus Professor of Nutrition in the University of London.

* To be published in Britain in July 1989 by Angus & Robertson, under the title *The Stone Age Health Programme*. Price will be £4.95, paperback only.

To the heart of the matter

Roy Porter

Vampires, Burial, and Death: Folklore and Reality. By Paul Barber. *Yale University Press*: 1988. Pp. 236. £14.95, \$30.

PAUL Barber's inquiry into vampires, fact and fiction, is a gem in the literature of debunking. We would all recognize a vampire if we saw one, that 'undead' soul who rises from the grave to gorge himself upon the blood of the living, thereby creating further generations of vampires: he would be tall, clad in a black cloak, and endowed with terrifyingly sharp fangs. But we would be wrong. For Barber reveals that this Count Dracula figure, so familiar from the movies, bears scant relation to the traditional 'revenant' who is so powerful a presence in Central and Eastern European folk memory. Examination of recorded sightings shows the vampire was quite different: above all, podgy and florid of complexion, and no mention of those teeth.

What are we to make of this traditional vampire? Folk wisdom had it that vampires were those who had died before their time (perhaps as suicides or murder victims) and, for that reason, were destined to find no rest. Their soul did not depart in peace, nor did their body decompose, and so they rose from their graves to haunt the living, whose blood they would suck for sustenance. They had to be captured, and then a timely stake through the heart or cremation would ensure that they would never walk again.

Of course, Barber is not suggesting that these supernaturalist beliefs should be taken seriously. What fascinates him above all are the attempts that intellectuals have made to provide 'scientific' explanations of well-attested 'vampire' phenomena, such as undecomposed bodies found outside their graves. Once popular, for instance, was the notion that the dead walk because they were never dead at all, but were unfortunates who had suffered premature burial, while in a deep coma, and who had regained consciousness and clawed their way out of their coffins. But such 'naturalistic' accounts are no less incredible than the old supernatural stories themselves: they are merely different versions of folklore.

A little elementary forensic pathology, Barber advises, will reveal the truth that lies at the heart (so to speak) of vampire myths. If the vampire is the body that refuses to decompose, that should not surprise us: as the gravedigger in *Hamlet* knew so well, under all sorts of climatic and soil conditions corpses do indeed



Wot no fangs? Bela Lugosi as Dracula approaches a prospective victim.

disintegrate very slowly. The presence of liquid blood around the mouth — supposedly the evidence of midnight snacks — is likewise easily explained: blood often does not coagulate. Moreover, the bloated condition of such corpses is not, of course, the result of gorging on blood, but the product of gaseous build-up. And it is the escape of these gases which accounts for the moans when the stake is finally driven through the heart. Floods and subsidence readily explain why buried corpses, or at least a hand or foot, rise to the surface from time to time.

In other words, corpses often do not look dead in the way that people expect them to, thus explaining the vampire myth that they are still alive. Belief in revenants is therefore a quite intelligent attempt to build a theory upon experience: pathological science is then needed to give the true explanation.

Two conclusions emerge from Barber's sleuth work. One is that the so-called 'scientific' explanations which scientists, past and present, often devise to explain, or explain away, the odd, the marvellous, the praeternatural and the paranormal, are often monuments of misplaced ingenuity, far-fetched in themselves and not even consonant with reported data. Science thus has a mythological folklore of its own.

The other underlines the terror that death strikes into the living. Through complex burial practices, mankind has always gone to immense trouble in disposing of the dead. Partly out of respect, indeed, but not least out of a need to propitiate the dead while at the same time immobilizing them. Fearing that the dead are those who are the cause of death, we have created secure burial mounds and coffins as much to protect the living from molestation by the dead as to honour the corpse. Even when safely buried, the suspicion lurks that they are not quiet, and hence they may need to be slain and slain again.

Barber's book is a convincing exercise in mental archaeology. It should be read alongside Pietro Camporesi's *The Incorruptible Flesh* (Cambridge University Press, 1988) which investigates a parallel phenomenon: the vast fascination which traditional Christianity has had for the flesh of saints and martyrs. We should not be complacent today about our own enlightenment in these matters, as public panic over the disposal of the bodies of AIDS victims amply shows. □

Roy Porter, Wellcome Institute for the History of Medicine, London, is currently at the William Andrews Clark Memorial Library, 2520 Cimarron Street, Los Angeles, California 90018, USA.

Strolling up the garden path

Igor Aleksander

Mind Children: The Future of Robot and Human Intelligence. By Hans Moravec. *Harvard University Press: 1988. Pp. 214. \$18.95. To be published in Britain on 9 December, £14.95.*

Mind Children is a frightening tale of the potential for the development of super-human 'intellect' in machines. The horror lies not so much in the actual content of the book, but in the fact that Hans Moravec suggests that a future in which the biological has been eradicated in favour of the technological is 'wonderful' and 'exciting'. To be fair, Moravec, as a sensible robot designer, gives himself trillions of years before seeing the realization of some of his predictions. But, he believes, this length of time can be bridged by 'transmigrating' a human being (consciousness, intellect, personality and all) into an immortal machine. Starting with the premise that in the near future there will be no essential human function, physical or mental, which cannot be artificially created, his argument follows a simple six-point progression.

Foundations

First, Moravec submits that the foundations of all the necessary techniques are already in place. Heading the list are Minsky's and Winograd's robot planning schemes, devised at the Massachusetts Institute of Technology in the late 1960s, which are said to be capable of understanding simple English-like sentences about worlds with simple objects in them. As further evidence, programs are said to exist that recognize sensory patterns and are capable of simple reasoning about them. Mobile robots of considerable agility have been designed, while learning and making general inferences has been shown to be possible with artificial neural networks. One problem, concedes the author, is that the technical knowledge has, as yet, not been married to engineering practice. This is needed to create a knowledge-gathering robot which can build up its own awareness of the world. Moravec accounts for this failing by suggesting that, because of the excessive cost, computer power has not reached the levels required. But given about 50 years, he says, the problem will have been overcome.

Consequently, the second pillar of Moravec's philosophy is that the acceleration in computer power per unit cost will continue. In particular, the ability of machines to absorb and classify masses of sensory data will increase hugely, and enable the mobile robot to explore its

surroundings in much the same way as a child. Moravec includes advances in biotechnology as a way of achieving some of the necessary packing densities of computing components.

The third step is to do with the symbiosis between man and machine. Improved sensors and displays will enable machines to work alongside human beings, allowing us to become helpful teachers to the robots. Devices such as 'magic glasses' and 'magic gloves' are invoked as existing early steps in making the computer provide sensory feedback to its human user. The gloves enable the user to 'feel' a simulated solid and so grasp and move objects on a screen. The glasses complete the illusion of actually being in some location that is merely stored in the memory of the machine.

The climax of this argument comes in the fourth step, where the world, driven by industrial competition, will see the development of the perfect computer receptacle into which the complete content of a human being's brain (hence intellect) can be transmitted. Moravec is aware that this notion can be demolished by the argument that such data would be useless to a machine that was not driven by its own needs to acquire it. He labels such non-believers as being besotted with some obscure viewpoint called 'the body-identity position', and urges them to shed this heresy and instead embrace 'the pattern-identity position'. This he clarifies as being the belief that 'I' am the process and pattern of what goes on in my head, and if that can be preserved in some other box 'I' become preserved in that box. In this way we shall all be able to be preserved forever, regenerating in a new box whenever the old one starts wearing out.

The fifth link in the progression suggests that a kind of independent, informational life can be created in the computer receptacles without the need to absorb human intellect. There is nothing to stop the receptacles from developing unlimited super intellects of their own. The evidence for this possibility is found in computer viruses (programs that destroy other programs unbeknown to the computer user); Conway's Life Game, which is played on a two-dimensional grid on which each point has simple rules about 'firing' or 'not firing' dependent on the firing of its nearest neighbours (limited patterns of grid activity survive and even shuffle around the grid to the surprise of the programmer); and Von Neumann's self-reproducing automata (programs transmitted from area to area on a grid of processors).

So we come to the sixth point, the *coup de grâce*, which defies the second law of thermodynamics. Because of the fall in noise with a fall in temperature, Moravec suggests that informational intelligence can remain constant as the Universe drifts towards absolute zero. What happens is

not clear, but by this point one does not care all that much.

Limitations

The trouble with predictions of this kind is that their persuasiveness lies in the disregard for the limitations of precisely those technological developments on which they are based. These deficits are hard to keep in mind as the reader becomes progressively blindfolded and is taken up the author's garden path. I would recommend that anyone who reads this book should immediately read another: Terry Winograd's and Fernando Flores's *Understanding Computers and Cognition* (Addison Wesley, 1986). This is a closely argued antithesis of Moravec's view, in which Winograd and Flores point out that 'intelligence' and 'understanding' as currently used in the context of computers may simply be metaphors that have no relevance to the creation of artificial intellect. From this standpoint, Moravec's reasoning collapses like a house of cards.

The ability of computers to 'understand' is shown to be merely a logical party trick (a juggling of pre-programmed phrases) by the very author (Winograd) of the programs on which Moravec bases the first step in his chain of reasoning.

Further, computer speed, cost and packing density have no bearing on machines becoming human-like. The missing ingredient is human consciousness itself, which is an integration of the experience of an individual's brain and body, closely tied to that body's make-up. This puts paid to Moravec's second and third pillars as it makes the 'body-identity position' seem sensible, and relegates the 'pattern-identity position' to science fiction. The rest falls into a crumpled heap: in particular, the Life Game and self-reproducing automata return to the realm of mathematical diversions, as relevant to philosophy as winning at Monopoly would be to becoming a real millionaire.

Moravec's book should be read, but perhaps not for the reason it was written. Where the author wishes to impress his audience with a vision of the way computer technology might lead to human immortality, readers should enjoy using no more than common sense to see if they can spot where the line between futures that can reasonably be predicted and those that belong to science fiction can be drawn. By not acknowledging the existence of views which differ from his own, Moravec makes no such judgements. For someone who is described as a serious scholar on the jacket of the book, that is a serious failing and one which vitiates his entire enterprise. □

Igor Aleksander is Professor of Neural Systems Engineering and Head of the Department of Electrical Engineering at Imperial College, University of London, Exhibition Road, London SW7 2BT, UK

Up among The Great and The Good

William Cooper

Monkeys, Men and Missiles: An Autobiography 1946–88. By Solly Zuckerman. Collins: 1988. Pp.498. £19.50. To be published in the United States by W.W. Norton.

LORD Zuckerman's autobiography is divided into two volumes of 400–500 pages apiece. The first, *From Apes to Warlords*, was published in 1978; *Monkeys, Men and Missiles* is the second, covering the years 1946–1988.

In the account of an unending series of important jobs Zuckerman has done between those dates, at the core of British affairs, 470 pages of text are scarcely one too many for the inside detail he is in a position to disclose. Every page is strewn with names of notable personages in Britain and in the United States, introduced formally by their full names and (excepting certain Prime Ministers and Presidents) subsequently referred to by their first names — Bob and Jerry, Patrick, Dickie, Ted and Dennis. . . . Zuckerman has known them all, and most of them, President Kennedy included, have called him 'Solly'. Among the jobs we have Professor of Anatomy at Birmingham University, Secretary of the Royal Zoological Society; military adviser with the techniques of operational research at his command, and authority, by experiment and observation, on the effects of nuclear blast (obliterating); Whitehall pundit for all seasons, wrestling with the organization of Britain's technical manpower and research and development; Chief Scientific Adviser to the Ministry of Defence; and finally Chief Scientific Adviser to the Government — maintaining even after retirement a 'perch' in the Cabinet Office. A fascinating story.

It makes for more fascinating reading, though, if the author of an autobiography has some of the gifts of a novelist — can make 'Patrick' spring to life as a man; can raise a dramatic spark in the absurd rebellion of the Old Guard on the Council of the Zoo; and, most importantly of all, reveals his own nature and personality. Without great evidence of this gift, the earlier part of the volume is pretty flat; one gets an impression of the author's personality chiefly by inference, insight into his nature practically none. However the latter part becomes a history, often day-by-day and sometimes hour-by-hour, of events whose central subject is nuclear missiles, a history so riveting that the absence of a novelist's gifts is irrelevant. One still wonders, all the same, how the unending series of jobs came about.

Here I'll air a theory of my own. It is that the Establishment, looking around for persons to fill important jobs, is incapable of holding more than one name in its corporate head at a time. For every job that comes up over a period, that one name comes up for it. Over the period in question it looks as if the one name that came up was Solly Zuckerman. Married in 1939 to the daughter of a former Viceroy



of India; elected in 1943 to a Fellowship of The Royal Society; his name was obviously eligible for the list of The Great and The Good, and had the fortuitous advantage of being very individual and endearingly tinged with the comical. Furthermore the way he went about his various jobs — a seemingly artless zest for his own manoeuvres and innocent delight in their success — had its appeal. For example:

The Council [of the Zoo] had agreed that I could ask the late Sir William Walton to compose a fanfare for the occasion. I had always envied the Royal Academy the Arthur Bliss fanfare at its annual dinners, and William was at least as famous a composer of fanfares. He obliged, and it was performed by the trumpeters of The Royal Military School of Music on the arrival of the Queen and Prince Philip. William entitled it 'The Roar of Lions' and dedicated it to me, calling me 'the lion of lions'.... The occasion went off splendidly, and I believe the Queen was almost as pleased as Charles Clore. [He was trying to raise money for the Zoo from Clore.]

Anyone who finds this passage a trifle

bumptious should remark that in his first volume of autobiography there is merely an oblique reference to his having been elected to the Royal Society; and in the second there is no mention of his receiving in 1965 the highest honour of all, the Order of Merit. And so far as his being on first-name terms with so many other people in the business was concerned, it was presumably to everyone's advantage that at any juncture, any crisis in affairs, he could suggest which buttons to press in the old boy network.

Coming to Zuckerman's main contribution to policy in his later years, one might argue that in posts such as Chief Scientific Adviser the incumbent is not called upon for the highest flights of scientific creativeness, but rather for a combination of commonsense and logic — which is indeed the identifying characteristic of Zuckerman's contributions. Commonsense has long seemed to me a sort of 'Layman's Science'. It enables one to explain and predict, over a range of human experiences in the real world, what happens and what is likely to happen, on the basis of accumulated observation of what has already happened and why. (This is not to say that commonsense doesn't have its lapses any more or less than science does; nor that all scientists are gifted with commonsense. Far from it! A case in point, the rock on which Zuckerman's friendship with the remarkably gifted scientist, J. D. Bernal, foundered, was the latter's lapses from, if not complete lack of, commonsense.)

Zuckerman's particular combination of commonsense and logic is nicely exemplified in what he calls "the inexorable law of R&D" *à propos* the arms race between East and West:

There is no finishing post to the race, of course, unless it were stopped by the ultimate disaster of an East-West war. The cost of rearmament or 'modernization' — as some now term the process to make it sound more palatable politically — cannot but go on rising at a rate greater than the rate of increase in the fund of resources that can be devoted to all public expenditure. This conclusion applies as much to the USA, and presumably to the USSR, as it does to Britain. The economic burden can on occasion be reduced in the short term by sales or by the application to civil industry of some useful technological development made in the course of a piece of defence R&D. That, however, occurs far more rarely than is thought.

It scarcely needs to be said that views based in commonsense plus logic are liable to be dubbed heterodox; they are not universally popular among politicians, are even less popular among War Lords and are frankly unpopular among 'weaponers'. Zuckerman makes no pre-

tence of always being in the right. "I couldn't have been more wrong", he says at one point. Nor of always claiming success. U Thant once sent him inscribed copies of reports he, Zuckerman, had written on arms control for the UN, about which he writes: "Despite their wide circulation I never did learn if any one of them had much effect on world opinion. The way the arms race has proceeded over the past decades I would think not".

However it is when he comes to nuclear weapons that most readers may well find his story at its most gripping. In a crucial passage he states another of his propositions:

The state of mutual strategic deterrence, upheld by the threat of retaliation with nuclear weapons, will remain valid until (1) a meaningful defence is achieved against both ballistic missiles and low-flying aircraft... or (2) a weapons delivery system is perfected which, for all practical purposes, could successfully carry out a surprise attack which destroyed so large a proportion of the other side's delivery system before they could be launched, as to make it impossible to make a retaliatory strike. [It was]... just technically possible, but immensely expensive in resources to devise an anti-ballistic missile capable of intercepting the simplest kind of incoming warhead.... But the balance is inevitably in favour of the attacker, and no Western scientist or technologist knows a way of providing an anti-ballistic missile defence of the major targets in his country. If the Russians know of one, we had no idea what it is.

Twenty five years later that remains my view, SDI or no SDI. The state of mutual deterrence is likely to remain valid indefinitely.

It is clear that Zuckerman is entirely convinced of the use of nuclear weapons in their role of deterrent: equally clear is that his confidence in their use in the cause of military strategy is nil. In this latter view he came of course into conflict with those weaponeers who, having 'perfected' their nuclear weapons, felt — and no doubt still feel, human nature being, alas, what it is — an irresistible urge to fire them off. He describes an encounter with Edward Teller, whom many people regard as the Evil Genius of The Bomb — in part responsible for its totally unnecessary use on Nagasaki, then fighting any moves whatsoever to ban test explosions, then shifting ground to propose the use of just little nuclear bombs on the battlefield, and finally as a last resort developing a case for exploiting nuclear bombs to make big holes in the ground. Zuckerman records the stormy meeting at which first he, and then Lord Mountbatten, denounced Teller's proclamations (about battlefield weapons) as dangerous military nonsense:

During a break in the proceedings, I was sitting chatting in the sunshine with some uniformed friends, when Teller came up to me, and, in his strange accent, said, 'I will not forgive you for that,' to which I replied that it would make not the slightest difference to me if he didn't. I knew that nothing could stop him. To this day

he continues to expound what to me is dangerous military and political nonsense.

It should not go without remark that Zuckerman in those days received a great deal of support and wise advice from Harold Macmillan, who was then Prime Minister. But in 1964 it was decided to 'reform' the loose components of the Ministry of Defence into a huge new 'super-department'. Thus:

... I nonetheless feared that the bureaucratic procedure he was suggesting could slow down action, particularly when minutes from my side of the Ministry were pored over by people who could not understand their scientific and technological significance.... Nor was I ready to surrender my right to say what I wanted, regardless of the views of the rest of the Ministry, and even those of my staff. This manifesto in favour of free speech was in the end to bring about my departure from the department.

Shortly afterwards Harold Wilson came to power, and wanted Zuckerman to go to the Foreign Office (with a peerage) as

Minister of State for Disarmament. He turned it down, wanting to remain a scientist rather than to become a politician. (Yet, some years later, he wondered if he could have achieved more for disarmament by accepting.) Instead he became Chief Scientific Adviser to the Government, lodged in the Cabinet Office. In this new role he held a succession of jobs over a wide range of scientific interest, environmental and so on. Yet a preoccupation with nuclear affairs has remained nearest to his heart. When the time came for retirement he accepted a peerage from the then Prime Minister, Edward Heath. And in 1982, still on his lasting theme, he wrote a book called *Nuclear Illusion and Reality* — which, one imagines, does not find entire favour with the now Prime Minister. □

William Cooper, Savile Club, 69 Brook Street, London W1Y 2ER, UK, is the pseudonym of the novelist Harry Hoff. From 1958 to 1972 he was personnel consultant to the United Kingdom Atomic Energy Authority.

Soviet star-turns

Desmond King-Hele

Race into Space: The Soviet Space Programme. By Brian Harvey. *Ellis Horwood: 1988. Pp. 381. £16.95, \$39.95.*

FUTURE historians may well rate the Soviet effort in space exploration as one of the greatest technological achievements of the twentieth century. The launching of Sputnik 1 in 1957 came as a great surprise to Western nations basking in the self-esteem engendered by their more advanced technology. That surprise has ripened into embarrassment as the sustained momentum of the Soviet programme has become evident.

If the journalistic cliché of a 'space race' is accepted, as in the title of Brian Harvey's book, it is a race in which one runner has already lapped all the rest: in the ten years from 1978 to 1987 there were 933 space launches by the Soviet Union (excluding Intercosmos), 146 by the United States (excluding international launches such as Intelsat), 24 by Japan and lesser numbers by others. Awkward statistics like these are conveniently ignored on the rare occasions when Soviet launches are reported in the West: we are usually told that the Soviets are beginning to catch up the Americans, and most people accept this perversion of reality, so awesome is the power of persistent propaganda.

Although there have been failures in the Soviet programme, the fact is that the Mir space-station with six docking points has been in operation since 1986: crews have been interchanged from time to time; there have been numerous visits

by short-stay cosmonauts; unmanned 'Progress' craft have called regularly to deliver supplies and remove rubbish; scientific experiments have been pursued aboard Mir and through modules like Qvant, attached at spare docking ports. Already there is an orbital community, with people and supplies brought in and out not by car and truck but by advanced vehicles appropriate to the environment. At present, living in space is neither spacious nor gracious: the living quarters are cramped, and most crews are all-male. But both of these limitations will gradually be relaxed if Mir succeeds in avoiding the ever-present threat of disaster (for example, from being struck by debris). We should then see larger space habitats with better-balanced populations. The Mir programme is also an impressive act of faith in the future good behaviour of humankind. 'Mir' means 'peace', and Mir needs peace, because this most expensive project can be ruined by any hostile power that chooses to detonate debris-creating weapons in orbit at the appropriate height.

Why has the Soviet Union devoted so much effort to space exploration, by comparison with the United States, which took the lead with the manned Moon landings in 1969–1972 and then seemed to lose interest? The easy answer is that the Soviet Union has had consistent government and long-term planning, while the United States suffers from the hot and cold turns of its electorate, and was lured into building a space shuttle for travelling between Earth and a space station that had already been cancelled. (Another is now being proposed, to give the shuttle a role.)

This answer has some validity, but does not tell us why the Soviets have been and

still are so keen. One of the best features of *Race into Space* is the attempt to explain the impetus in historical terms. The inspiration came from Konstantin Tsiolkovsky (1857–1935) who, in books such as *A Dream of Earth and Sky* (1895), foresaw nearly all of the events that have actually occurred. He advocated liquid-fuel engines, the staging of rockets, solar power and even stellar ships with closed ecocycles. Surprising as it may seem to those living in Western countries, where rocket enthusiasts were scorned until after 1940, Tsiolkovsky was encouraged by the state; he was given a pension in 1918, and

There is very little about the photographic reconnaissance satellites, of which nearly a thousand have been launched, or about the many satellites for ocean surveillance, electronic ferreting and early warning. The systems of satellites for communications, meteorology and navigation are only glanced at, and there is scarcely a mention of the numerous spacecraft carrying scientific experiments.

The strength of the book is in its detailed accounts of the out-and-back journeys of the cosmonauts, all of whom are mentioned by name. Not all have been Soviet citizens: they have come from

many different countries, ranging alphabetically from Bulgaria and Cuba to Syria and Vietnam (with Afghanistan as a recent addition). With so many flights to be covered, the details of the crews may seem almost as boring as a roster of bus drivers leaving and returning to their depot. Harvey recognizes this problem and tries to enliven the text with charac-

ter-sketches of some of the cosmonauts and overdone dramatizations. For example, there is the story of the US astronauts arriving by airliner at 'Star City': "out jumped the three Americans", we are told, and after this improbable gymnastic feat they drive in a minibus past "endless avenues" in this city surrounded by desert, "endless, flat and arid", with a river running through it. Wet or dry, the style is not flat, and the adjectives are as endless as the avenues or the desert. At the other extreme, the crew of Voshkod 2, landing at night, had to endure "frosty breath from lurking shadowy forms" (these are wolves, not the welcoming party).

There are plenty of photographs, not of wolves, but chiefly of cosmonauts, who all look remarkably similar in space suits, especially as the quality of the photographs is poor. Also there are no diagrams in the main text, and the author is left with the unenviable task of trying to describe in words the complex design of Mir. All in all, the book does provide a useful, detailed account of Soviet manned spaceflight and lunar and planetary exploration, though the style may deter some would-be readers and minor inaccuracies are numerous. □

Desmond King-Hele, 3 Tor Road, Farnham, Surrey GU9 7BX, UK, was formerly a Deputy Chief Scientific Officer at the Royal Aircraft Establishment, Farnborough. Among other books he is author of *Satellites and Scientific Research* (1960) and *Satellite Orbits in an Atmosphere* (1987).

Arp's astronomical exotica

Michael Rowan-Robinson

Quasars, Redshifts and Controversies. By Halton Arp. *Interstellar Media*, 2153 Russell Street, Berkeley, California 94705/ Cambridge University Press: 1988. Pp.198. \$22.20, £12.50.

In 1966 Halton Arp compiled one of the most fascinating catalogues of modern astronomy, the *Arp Atlas of Peculiar and Interacting Galaxies*. For nearly 20 years the remarkable objects in this catalogue were for the most part ignored by astronomers, who continued to think of galaxies as basically regular-shaped spirals or ellipticals. Now, in the past five years, interactions and mergers between pairs of galaxies have leapt to the forefront of the astronomical consciousness, partly because of the growth in sophistication of computer simulations of the dynamical interactions between galaxies, and partly because a higher than expected proportion of luminous infrared galaxies appear to be due to interactions or mergers.

But Arp will not be able to enjoy this tribute to his perspicacity. For as he was preparing his *Atlas* for publication, he noticed that some of the sources in the *Third Cambridge Survey of Radio Sources* appeared to lie in pairs on opposite sides of some of his peculiar galaxies. Because the recession velocities of the radio-galaxies and quasars with which these sources were identified were in general enormously larger than those of the 'associated' peculiar galaxies, not only must the Hubble velocity-distance law break down in some circumstances but galaxies must be capable of spewing out whole galaxies and quasars.

This discovery took Arp down an extraordinary trail of quasars associated with nearby galaxies, companion galaxies with excess redshift, chains of galaxies with discordant redshifts, to quasars as members of the Local Group of galaxies. Because quasars were at that time so novel and incomprehensible many astronomers, myself included, were at first willing to keep an open mind about Arp's associations and to follow him at least a little way down the trail. Could the quasars, for example, like the nebulae before them, comprise more than one completely distinct kind of object?

Gradually, the possibility of there being any serious basis to Arp's claims dissolved. Quasars were found to be situated in galaxies and were often found to be in groups or clusters of galaxies which had the same redshift as the quasar. Statistically controlled surveys for quasars showed them to be distributed homogeneously



On top of the world — the cosmonauts involved in the Salyut 1 mission. The picture is reproduced from *The Soviet Manned Space Programme* by Phillip Clark, Publisher is Salamander, price is £14.95.

a meeting of the Academy of Sciences was held in his honour in 1932. His home was to become a place of pilgrimage. Scientific achievement (unsullied by the profit motive) was held dear by the Soviet government in the 1920s and early 1930s, when the engineer Frederick Tsander (1887–1933) constructed liquid-fuel rockets, the author Yuri Kondratyuk foresaw the future course of space exploration, and research institutes for jet propulsion and aerodynamics were established. "By the mid 1920s, all the major components of the Soviet space programme were in place", Harvey remarks, and it seems a defensible claim. But the programme ran into the sands, because of policy changes and Stalin's purges — among those imprisoned was Sergei Korolov (1906–1966), later the 'chief designer' of space launchers. After the Second World War, when the German V2 showed everyone what could have been done, Soviet rocket development went ahead strongly. The desire for military rockets was the driving force, but their possible use for space launching was quickly recognized, and exploited by Khrushchev from 1954 onwards.

After this historical chapter, *Race into Space* becomes a detailed narrative, seen always in terms of rivalry with the United States, of Soviet manned spaceflight and lunar and planetary exploration. Only a few pages are devoted to the many other types of satellites that make up the great majority of the Soviet space launches.

over the sky. The Hubble law has been confirmed with ever-improving precision.

Today, Arp is left bitter and combative, with only a handful of astronomers such as Fred Hoyle and Geoff Burbidge taking him seriously. In this book he collects together all the best examples of his wondrous phenomena and an unimpressive and unconvincing assembly it makes. The 'lines' and 'concentrations' of quasars appear no more unusual than would be expected by chance, given that Arp has worked extremely hard to find these examples. The high-redshift galaxies near foreground objects always look just like more distant, background objects. Every so often Arp stumbles on some real phenomenon, such as the interactions of his 1966 *Atlas* or the optical jet of M87, but even then his explanation of it always has to be totally bizarre.

At one point Arp resorts to a rather sad piece of sophistry. Discussing the galaxies of Stephan's Quintet he remarks that a supernova had been discovered in 1971 in NGC 7319, the galaxy in the group which

has an anomalously high redshift. Rather than accept that the faintness of the supernova supported a higher distance for the high-redshift galaxy, Arp concludes that: "stars in the anomalously redshifted systems simply do not get as bright as in galaxies of lower intrinsic redshift". This is sad because in 1961 Arp was the first to apply the Baade-Wesselink expansion parallax method to supernovae, and so could be called the father of the supernova distance method.

Arp claims that he has been harshly treated and kept off telescopes and out of astronomical journals. Yet he has managed to publish over 40 papers on his exotica over the past 20 years. He was president of the Astronomical Society of the Pacific from 1980 to 1983, and has received a sprinkling of prizes and awards. This book illustrates the maxim that an obsession can ruin your whole life. □

Michael Rowan-Robinson is Professor of Astrophysics in the Astronomy Unit, Queen Mary College, Mile End Road, London E1 4NS, UK.

The universal rover

Owen Gingerich

The World Within the World. By John D. Barrow. Clarendon: 1988. Pp. 398. £20, \$24.95.

How does the uniformity of nature depend on the quantization of energy levels? I can only admire a 'popularization' that has the ingenuity to ask such a question. Here John Barrow, an astronomer at the University of Sussex, not only asks the question, but provides a notably clear answer: if the electrons orbiting protons could have any arbitrary speeds, then each hydrogen atom would be different, and "the uniformity of Nature an unrealized idealization".

In this wonderfully wide-ranging excursion, Barrow has brought a freshness of descriptive metaphor to some of the more arcane topics of modern physical science such as the Einstein-Podolsky-Rosen paradox, gauge theories, chaos, superstrings and the inflationary universe. Barrow explores not only a wide gamut of topics in the zone between physics and astronomy, but he wrestles with their philosophical contexts as well. His "world within the world" is the submicroscopic quantum framework within the large-scale cosmos, so strangely incompatible with the relativistic picture of the vast astronomical Universe and yet so essential to its understanding. As he describes the march of science, he circles around a far more difficult question: are there really laws of nature out there to be discovered, or are we just imposing our own human

and sometimes incompatible structures on a deeply unknowable Universe? Is there a constructed world within an ultimately unfathomable 'real' world?

Barrow's roving exploration is filled with a rich but comfortable level of erudition. The sources of the ever-apt quotations that open the sections range from Isaac Newton, Michael Faraday and Richard Feynman, to Mark Twain, Ambrose Bierce and Groucho Marx. The reader can discover that the Apostle Paul quoted the paradox of Epimenides (about a Cretan stating that all Cretans are liars), or contemplate Barrow's anti-popperian critique: "notions like falsification, verification, and operationalism seem like gloves that ill-fit the hand of Nature, however fashionable they may appear on the hands of idealized scientists".

The World Within the World is a singularly honest account compared with the sleight of hand that characterizes Stephen Hawking's best-selling *A Brief*

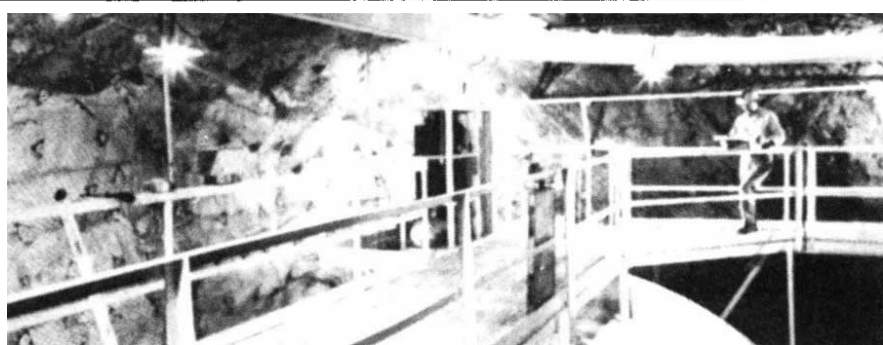
History of Time. Hawking applies a coordinate transformation to the singularity of the Big Bang, expands the time coordinate back to infinity, and glibly announces an originless Universe, asking "What place, then, for a creator?". Barrow, in contrast, discusses the question of singularities at some length, and concludes that until we better understand the nature of time, we will really not know how to answer those who ask whether the Universe is of finite or infinite age.

Although it achieves a sophisticated level of popularization, Barrow's *World* is nevertheless an uneven book. The conversation between a mathematical intuitionist and a platonist is a confusing detour, and Barrow's road map of philosophical routes is too terse to set his readers onto clear paths. In his assessment of Thomas Kuhn's paradigmatic structure of scientific explanations, and elsewhere, he subtly suggests that the highest philosophical calling of a physicist is to be a realist. Yet his sympathetic discussion of the 'many worlds' interpretation of quantum mechanics, his description of multiple dimensions of the Universe, and his emphasis on metaphysical and aesthetic criteria as guides to theoretical speculation seem to make him far more the idealist.

Indeed, after Barrow's rewarding journey to the limits of space and time, his conclusion is not all that unexpected:

[these are] extraordinary problems, and they possess extraordinary solutions which will require extraordinary methods to coax from the Universe. If our methods ultimately fail, then any boundary between fundamental science and metaphysical theology will become increasingly difficult to draw. Sight must give way to faith. Confronted with an emotionally satisfying mathematical scheme which is 'simple' enough to command universal assent, but esoteric enough to admit no means of experimental test then, closeted within our world within the world, we might simply have to believe it. □

Owen Gingerich is a Professor in the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.



Deep thinking — a view of the solar neutrino experiment located at a depth of 4,850 feet in the Homestake gold mine. The picture is reproduced from Particle Physics: A Los Alamos Primer edited by N.C. Cooper and G.B. West, published earlier this year by Cambridge University Press. Price is hbk £30, \$49.50; pbk £9.95, \$14.95.

Irregular tricks of the trade

Ian Stewart

The Science of Fractal Images. Edited by Heinz-Otto Peitgen and Dietmar Saupe. Springer-Verlag: 1988. Pp.312. DM68, £23, \$34.

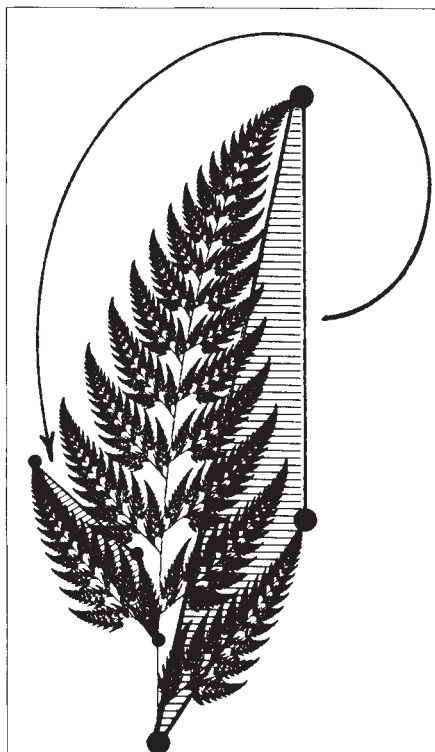
THE theory of fractals, the brainchild of Benoit Mandelbrot, has been responsible for some of the most arresting and memorable images in modern science, vividly attesting that mathematics does indeed possess a deep inner beauty. The fake landscapes of Richard Voss and the dramatic complexity of the Mandelbrot set displayed by Heinz-Otto Peitgen and Peter Richter have introduced an entire generation to the mathematics of irregular forms. Yet in some ways the theory of fractals is a subject in search of its own identity — it is a collection of examples from innumerable disciplines, linked by a common point of view, rather than an organized body of knowledge and technique.

For example, one of the most basic concepts of the theory is that of the Hausdorff dimension, a generalization of the usual notion of dimension which need not be a whole number, and which quantifies the degree of irregularity of a fractal. There exist no satisfactory general methods, other than numerical approximation, for calculating the Hausdorff dimension of most fractals, even those defined by very specific mathematical processes. It is a bit like the early days of calculus, when the only way to differentiate or integrate functions was to approximate them by a polynomial and differentiate or integrate *that*. Over the centuries the techniques of calculus have improved, and their ramifications occupy the main part of today's mathematics. The early work was dramatic, despite the lack of technique, and the same is true of the theory of fractals.

The Science of Fractal Images provides convincing evidence that those working with fractals are not resting on their laurels. In particular they are repaying a debt by stimulating new ideas in mathematics and computer graphics. Anyone interested in the internal logic of the fractal world will find the book irresistible even though it is a multi-author work with a variable level of technical difficulty. There are more than enough decorative pictures, but now their creators reveal some of the tricks of the trade. The volume originated in a SIGGRAPH course on computer graphics in 1987. At its heart is a collection of algorithms for the efficient generation and analysis of fractals.

Benoit Mandelbrot opens the tale with an essay on some of the people and events behind his theory. The story improves with the telling, and is still fresh. The five main contributions demonstrate the breadth, as well as the depth, of the mathematics involved. Richard Voss describes how to simulate natural forms, such as mountains, lakes and clouds, using probabilistic methods, and applies them to the analysis of variability of musical sounds. Dietmar Saupe enters more deeply into the underlying mathematics of fractals and the simulation of random noise with specified statistical characteristics.

Robert Devaney sets out some links



Order in chaos. Define four simple transformations of the plane, one of which is indicated by the arrow. Take an initial point, select one of these four transformations at random and apply it to that point: repeat indefinitely. The result here, reproduced from Michael Barnsley's contribution to the book, is a simulation of the black spleenwort fern.

between fractals and dynamics: the theory of chaos is surely one of the main contributions of mathematical thought to science during the latter part of this century. The central icon in the theory of fractals is the Mandelbrot set, a strange, squat form like a fat cat, budding kittens like a cactus and trailing tendrils resembling forked lightning. The Mandelbrot set combines in a single figure the behaviour of an entire complex plane's worth of equally striking forms, known as Julia sets. These sets are associated with complex functions $z \rightarrow z^2 + c$, for complex constants c . Devaney considers analogous sets for other func-

tions, especially transcendental functions such as the sine or the exponential. He has made colour movies of these sets, and some stills are included here. New mathematical phenomena, notably the tendency of Julia sets to 'explode', were discovered as a result.

Heinz-Otto Peitgen weighs in with a whole series of efficient methods for plotting Julia and Mandelbrot sets, and with an analysis of some surprising effects associated with Newton's method for the numerical approximation of solutions to equations. An appendix by Yuval Fisher develops these methods in more detail. A particularly striking algorithm generates the Mandelbrot set, or magnified sections of it, quickly enough to run on a desk-top PC. The algorithm is based on new geometrical ideas of John Milnor and William Thurston, and uses a recursive procedure to find circular disks that lie entirely outside, or entirely inside, the Mandelbrot set. The delicate and infinitely filigreed boundary rapidly emerges, sandwiched between the regions paved by these disks. The 'obvious' algorithm, based on the original definition of the Mandelbrot set, would take hours to run on a PC: this one takes a few minutes. To achieve this kind of speed previously required a dedicated machine with about 40 transputers running in parallel, or a supercomputer. Thought can be as effective as raw computational power.

Finally, Michael Barnsley describes how to design fractals to order by playing the 'chaos game'. Repeatedly apply one of a small set of transformations, selected at random, to some chosen point and the result is not a random mess but a highly structured fractal. A simple method for choosing the transformations permits almost any shape to be generated in this manner: plant-like forms — ferns, trees, leaves — appear to work especially well. Could this method be the basis of a system for graphical data compression, speeding the transmission of images and making their storage more efficient?

'Prehistory' aside, fractals have been around for about 20 years, and only took off during the past ten. During that time the theory has begun to embed itself in the mainstream of science and mathematics, as a natural tool for the description and analysis of irregular structure. Indeed certain kinds of irregularity are now perceived more as a new type of regularity, superficially random but with a precise internal organization. This is the enduring contribution of fractals: they give us the ability to see a new kind of pattern in the world around us. *The Science of Fractal Images* shows that that process continues to gain momentum and to refresh our ways of thinking. □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

But of course . . .

Roger Pynn

Fractals. By Jens Feder. *Plenum: 1988.* Pp. 283. \$49.50.

IN reading this book, I kept wondering for whom it had been intended. Not for the idly curious who are searching for an intuitive grasp of the concept of fractals: Benoit Mandelbrot's *The Fractal Geometry of Nature* provides that so much more elegantly. Nor for the mathematician, who can be left with no lingering doubts: although I cannot direct such purists, I know that Feder's book would not satisfy them. Finally, it dawned. If I were teaching a course on fractals to any group of natural scientists, this would be the course book. Of course, such is the route that Professor Feder has followed, and had I paid more attention to his preface I would have realized as much.

With the exception of the chapter on multifractals, which is confusing and undirected, the book is pedagogically sound. It gathers together in one place the various definitions of fractal dimension which are used in analysing experimental data and provides ample references to their use. The chapter on percolation is a clear discussion of this phenomenon and its relation to fractal concepts, and includes several new results which have helped to unify the subject. I particularly enjoyed the account of fractal records in time and rescaled-range analysis, which I have not

previously seen explained with such clarity. The link between this subject and studies of brownian motion is well described, and is illustrated by a series of graphs and pictures obtained from simulations programmed on a personal computer. Essential differences between self-similarity and self-affinity are dealt with approachably and succinctly. Several algorithms for the generation of fractal surfaces, landscapes and clouds are discussed, but Feder's book is intended to introduce this topic rather than to celebrate a new art form.

A great number of the examples used in the book would make excellent research or teaching projects for students. Perhaps, indeed, that is how they were born. The means required to study these examples are modest: a PC, a few purchases at a local hardware store and access to a public library would probably allow an interested student to generate enough data for another monograph within a year. There are, too, lots of references to the recent literature, providing the reader with a window on more than 200 other works.

Like Feder, I believe that Mandelbrot's concept of fractal geometry provides a sturdy framework for the analysis of natural phenomena in diverse fields of science. If that view continues to spread, it will not be long before the study of fractals is a mandatory part of university curricula. The book to support such a course is already available. This is it. □

Roger Pynn is at the Los Alamos National Laboratory, MS H805, Los Alamos, New Mexico 87545, USA.

Modern virtues

Richard S. Westfall

The How and the Why: An Essay on the Origins and Development of Physical Theory. By David Park. *Princeton University Press: 1988.* Pp. 459. \$35.

WELL into *The How and the Why*, I found that I was becoming annoyed with it. The subtitle appears to offer a book in the history of physics (or of physical theory), and the contents, beginning with the Greeks and proceeding chronologically through antiquity, the Middle Ages, and the scientific revolution towards contemporary physics, confirm that impression.

In all honesty, I was not greatly impressed. The chapters on the Greeks are disjointed collections of short essays about individual philosophers. There is no coherent account of the considerable mediaeval enterprise in natural philosophy. When I got to the scientific revolution, the period in which I claim some expertise, I found numerous errors, and my ill humour was by that time spilling over

freely into my notes. Then I arrived, about two-thirds of the way through the book, at the twentieth century and Park's discussion of contemporary physical theory. My irritation evaporated more quickly and more completely than dew on a summer day in the Middle West.

Never mind the subtitle; *The How and the Why* is primarily about the revolution in physical theory in the twentieth century. There are two different ways in which one can describe Park's discussion. On the one hand, it is popularization, for it proceeds in prose rather than in mathematics, with the small amount of mathematics that is included serving more to illustrate the argument than to carry it. When I say popularization, however, understand popularization on the level of Eddington's *Nature of the Physical World*. The reader who knows nothing about contemporary physics and expects to find it boiled down into digestible pap should go elsewhere. Serious students, who are not themselves physicists, but who want a discussion of one of the most important intellectual achievements of our century in terms as sophisticated as they can be without the mathematics, need look no further.

More than popularization, however, Park's book is also aimed at philosophically inclined physicists who want to stand back for a moment and contemplate the forest instead of the trees. In this respect *The How and the Why* considers not so much the content of contemporary physics as the nature of physical theory. I cannot claim expertise in modern physics, and the definitive word on Park's book will have to be delivered by others. But I find it difficult to believe that they will not be favourably impressed.

I began with words of criticism, so I owe it to candour to indicate briefly their foundation. I confine myself to Newton. Park asserts that "no letter *t* or its equivalent ever appears [in the *Principia*]. The only numbers in the *Principia* relate to distances; everything about time is implicit" (p. 232). I invite him to look again at Prop. VI, which enters into the argument of most of the rest of Book I. A bit later (p. 248) Park appears to state that Newton's work contains only geometrical diagrams and the waving of hands but no proofs. I have no idea what he can mean by those words, but it is hard to believe that anyone who says as much has seriously studied the *Principia*. I could multiply instances concerned with Newton and other figures of the scientific revolution at some length, and in a book that offers itself as an essay on the origins and development of physical theory, this constitutes a considerable fault. But I have made the point and wish to go no further. Everything in the book builds toward the final third, and one should not obscure its achievement by cavilling over subordinate matters.

In the climactic section of *The How and the Why*, extended accounts of relativity and quantum mechanics and of the evidence on which they rest lead on to the most recent endeavours in physical theory, such as superstrings, and to related work in cosmology. Throughout, Park concentrates on general themes, such as simplicity, symmetry and determinism, which were also central to his historical account of physical theory. One who is not a physicist cannot comment definitively on these chapters, and beyond the inherent complexity of the subject is Park's inclusion of the latest developments in contemporary physical theory, which will be new to the vast majority who are not themselves theoretical physicists. I am not a complete stranger to literature of this genre, and wish to say that I have never encountered more enlightening explications of these difficult issues. I wish Park had got the historical details a bit more straight. But I happily settle for the extraordinarily lucid and penetrating analyses of contemporary physical theory. □

Richard S. Westfall is a Professor in the Department of History and Philosophy of Science, 130 Goodbody Hall, Indiana University, Bloomington, Indiana 47401, USA.

Six views of the Universe

Harry Shipman and Sharon Juska

Creation: The Story of the Origin and Evolution of the Universe. By Barry Parker. Plenum:1988. Pp.297. \$22.95.

The Dark Side of the Universe. By James Trefil. Scribners:1988. Pp.197. \$17.95.

The Big Bang, revised and updated edition. By Joseph Silk. W. H. Freeman:1988. Pp.485. Hbk £19.95, \$24.95; pbk £9.95, \$13.95.

Gravity's Lens: Views of the New Cosmology. By Nathan Cohen. Wiley:1988. Pp.237. \$19.95, £12.95.

The Supernova Story. By Laurence A. Marschall. Plenum:1988. Pp.296. \$22.95.

Relatively Speaking: Relativity, Black Holes, and the Fate of the Universe. By Eric Chaisson. W. W. Norton:1988. Pp.254. \$18.95.

FOR astronomers, the past few years have been an exciting time. What seemed to be a virtually complete picture of the evolution of the Universe in the late 1970s has been transformed by the connection between particle physics and cosmology. This 'inner space/outer space' connection has spawned many exciting ideas, such as the hypothetical 'dark matter' and the origin of large-scale structure in the Universe. When an obscure blue star called Sanduleak-69 202 exploded in the Large Magellanic cloud, its violent death throes sent shock waves through its immediate vicinity, lighting up neighbouring gas clouds. Some 160,000 years later, this supernova sent shock waves through a community of astronomers living far away on Earth, dramatically confirming our ideas on how massive stars die in this most spectacular of celestial events. And then there remains a curious galaxy of astronomical beasts such as black holes, energetically interacting binary systems like SS 433 (the only star which had a finding chart shown on the *Saturday Night Live* television show), and gravitational lenses.

There are far more popular books on astronomy than on other scientific subjects, and still they come. The half dozen reviewed here, all addressed to the proverbial 'layperson', largely succeed in communicating astronomy's excitement both to scientists in other fields and to the public at large. All that is asked of the reader is a tolerance for astronomically large or small numbers and for a few strange uses of language. For instance, astrophysicists use the word 'inflation' to describe an abrupt increase in the size of the Universe which occurred 10^{-34} seconds after the Big Bang, not rising prices.

Four of the books deal to a greater or lesser degree with cosmology, the evolu-

tion of the Universe. The emphasis on 'classical', well-established research and on contemporary work which may be more speculative varies. The most classical of the four is Barry Parker's *Creation*, a well-conducted tour through the standard Big Bang theory spiced with just a bit of dark matter and such. It is now quite well established that the Universe did have a hot beginning some 10 to 20 billion years ago, and it is this idea rather than contemporary embellishments that Parker primarily concentrates on.

Parker is particularly good at bringing scientists to life. His descriptions of Hubble's brief forays into boxing, Gamow's impish sense of humour and boundless energy, Mike Turner's two-



Catch a falling star — the Leonid meteor shower of November 1833, which appears every 33 years.

year career as a car mechanic and other personal vignettes are marvellous. The illustrations complement the text well. Parker includes, for instance, the splendid picture of Gamow emerging from a bottle of Cointreau labelled as "YLEM" — Gamow's term for the primordial material.

Occasional organizational slips (black holes are stuffed in for no good reason) and infrequent lapses into leaden prose detract in only a minor way from what is generally an excellent treatment. Authors of books such as this always have a problem in distilling our understanding of particle physics into one chapter, but Parker meets the challenge, presenting a picture which was clear to the one of us who had never encountered particle physics before

(S.J.) and was still illuminating to the other (H.S.).

James Trefil covers much the same ground as Parker, but with greater emphasis on current controversies. Trefil's sceptical, even irreverent attitude towards currently fashionable cosmology is a magnificent antidote to the information-dump style of science textbooks and even some popular books. Trefil even includes a little table where he estimates the probability that various ideas are correct. He gives the basic Big Bang theory a 99 per cent plus chance of being right, while hypothetical yet popular superstrings are rated at 20 per cent. The main topic dealt with in the book is 'dark matter', a sea of particles of currently unknown nature which, according to current prejudices, should make up at least 90 per cent of the contents of the Universe and play a key role in the formation of galaxies. There is, too, a thought-provoking section on how understanding the evolution of the Universe does not mar our appreciation of its beauty, a defence of the scientific approach to cosmology.

Trefil's sense of humour comes through in chapter titles such as "Five Reasons Why Galaxies Can't Exist". At one point, in the middle of a difficult discussion of supersymmetry, he asks the reader directly, "are you *sure* you don't want to skip ahead to the summary on page 255?". Those who find the 10- and 26-dimensional worlds postulated by some modern particle theorists incomprehensible, incredible or both can find some comfort here.

Joseph Silk's *The Big Bang*, originally published in 1980, has now emerged in revised and updated form. This book, by far the longest and most intellectually challenging of the six reviewed here, is not really for the same audience as the others, even though all equations are consigned to a 40-page mathematical appendix. People with no science background can certainly get a great deal from much of the text, and can appreciate the more thorough treatment of the philosophical and mathematical foundations of the subject. But it is not always easy going, and some parts such as the three-chapter section on galaxy formation, evolution and clustering may be out of reach of strictly lay readers. This book is drier and less personal than the others; for example, the story of how Alan Guth discovered inflation, a science classic comparable to the story of Archimedes running naked through the streets, crying "Eureka", is all subsumed in the two words "brilliant insight".

But length and depth also have their virtues. For example, the history of cosmological thought is carried back as far as Pythagoras. Silk treats the 'cosmological principle', the assumption that the Universe will look the same in all directions from all points within it, in considerable

From *The Big Bang*

detail, its copernican foundations included. All of the modern stuff is there, among it a reasonably detailed description of why inflation is so compelling. Silk even has a chapter on 'alternative cosmologies', even though most of them are straw edifices which are easily pulled down. A reader of Silk's book will confront rather than avoid the peculiar properties of curved space which are characteristic of some cosmological models.

Nathan Cohen's *Gravity's Lens* falls short of the high standards of the others. His writing is too confusing, and he introduces too many terms which are underlined as though the reader has to remember them. You can understand how a particular type of star called a Cepheid variable can be used to measure the distance to galaxies without memorizing definitions of inverse square law, apparent magnitude, absolute magnitude and distance modulus — all of which are introduced in two pages! Too many factual slips mar the book (for example, quasars were identified as optical counterparts to radio sources and not, as Cohen says, as variable stars; it is the BL Lac objects which were first recognized as variables). On the good side, this book includes all the weird beasts that inhabit the astrophysical zoo and make the Universe such an interesting place. The treatment of gravitational lenses, which includes an account of the role of gadflies Jeno Barnothy and Fritz Zwicky in anticipating their discovery by decades, is decent enough.

Larry Marschall's supernova book is a superb piece of writing; it reads almost as easily as a good novel. Marschall's vivid descriptions bring astronomy to life. For instance, he reports that the radio telescope through which pulsars were discovered looked "like a giant colony of clothes lines waiting for wash day". He resists the temptation to pay undue attention to the recent supernova, and presents a balanced look at the whole field rather than a journalistic account of the life and times of supernova 1987A. He begins with a general treatment of astronomy, showing how supernovae fit into a bigger picture, and includes a very complete account of historical observations of these celestial explosions, in which stars suddenly blaze forth into the night-time, and even the day-time, sky, then fade away. The book is remarkably up to date — there is even a complete discussion of the mysterious spot and the light echoes which were discovered in mid-1988.

Chaisson's brief treatment of relativistic astrophysics is less successful, though one chapter, "Einstein, the Man", is pretty good. It is a slim collection of 30-second sound bites and pictures loosely organized about Einstein's theory of relativity and its astronomical consequences. Over-statements abound. For example, Chaisson writes that "... people from all segments

of society seek to know whether the expansion of the Universe will continue for ever". He can't be serious. This question does not even count as a hot one for cosmologists in the 1980s, much less the average person; the inflationary model basically provides a positive answer. Indeed, the treatment of cosmology is rather dated, with little attention being paid to dark matter and the large-scale structure; it reads as though most of it were written ten years ago.

It seems to have become the fashion in books such as these to have few or no references either to other books or to the research literature. Although a popular

book does not demand the historian's detailed attention to footnotes and so on, some indications of further reading, or of where one can follow up a particular topic, are surely essential. Trefil includes a few passing references (but only to his own books), while Parker and Cohen have none at all. So here the plaudits go to Chaisson, Marschall and Silk, three authors who have taken the trouble to ensure that their books contain useful bibliographies. □

Harry Shipman is Professor of Physics and Astronomy and Sharon Juska is an undergraduate in English at the University of Delaware, Newark, Delaware 19716, USA.

Mathland on \$17.95

C. J. S. Clarke

The Mathematical Tourist: Snapshots of Modern Mathematics. By Ivars Peterson. W. H. Freeman:1988. Pp.240. \$17.95, £15.95.

WE RESIDENTS have an ambivalent attitude to tourists: pleased that they are interested in the locality, irritated when they trample on the flowers. But Ivars Peterson behaves himself impeccably in his tour of the world of mathematics — one would almost think him a native — and his modestly subtitled collection of "snapshots" would serve as a good guidebook to the territory.

The emphasis is mostly on present-day culture rather than historical monuments, concentrating entirely on growth points in pure mathematics. Dramatic achievements of the past, even the recent past, such as the use of model theory in studying decidability in logic, are not explicitly included, nor are unsolved questions where there is little progress such as the Riemann hypothesis. The Poincaré conjecture, however, is given its proper place, along with the dramatic human interest of the failed Régé-Rourke attempt at proof.

The result is a very representative survey of those topics that have recently gained prominence. It should be accessible to quite a broad population, including the many readers who have been inoculated against all forms of mathematical reasoning by exposure to a de-natured form of it in their youth. Peterson succeeds in conveying a surprisingly good feel for the content of mathematics without, most of the time, actually doing any.

There is an inevitable bias towards the more spectacular and less abstract areas. Thus the account of the Mandelbrot set is one of the most successful parts of the book; but it is well placed in context, linked with the Julia sets and fully defined, and one forgives the few occasions when Peterson gets carried away by the graphics

("iridescent dragons clawing their own tails ... and geometric jets spewing colored streams into still, black basins ..."). At the other extreme of conceptual difficulty, he introduces the discovery of the non-standard differential structure on $\mathbb{R}^1$ by Donaldson. It is doubtful whether the concept of differential topology (as opposed to topology or geometry) will really get across to those not already familiar with it, but a valiant attempt is made.

A recurring theme is the use of computers. Indeed, if there is any 'message' in the book it is that computation is changing the nature of mathematics. Peterson shows computers as vital to the exploration of a variety of areas by providing visual representations of mathematical objects and so suggesting both new conjectures and ways of proving them; computers as themselves sources of mathematical problems, concerning the encryption of data, protocols for computer communication or the whole area of computational complexity; and computers as playing an actual part in the proof of theorems (though the four-colour theorem remains the only major instance of this).

There are frequent mentions of applications, exports from mathematics to other lands, but these tend to be unconvincing, the applications being of doubtful utility and seeming only palely to reflect the fervour of the research front. Many astronomers, for example, would question the assertion that "Astrophysicists have now confirmed that ... mass within the universe is distributed throughout space like a three-dimensional Cantor set", while the applications of knot theory to DNA structures involve little mathematics that was not available in the nineteenth century. The truth is that the real value of mathematics is not to be found in the latest Big Theorems, but in the whole integrated body of techniques and ideas that are needed to prove the theorems, and which they in turn give rise to. □

C. J. S. Clarke is a Professor in the Department of Mathematics, University of Southampton, Southampton SO9 5NH, UK.